Common consent

The distribution of human cell lines used in research should not be hindered by restrictions from donors.

Scientists rightly expect to have access to the materials described in a published research paper, so that they can verify the results and build on the findings. But, as the Corrigendum on page 1044 of this issue shows, the act of research and the ethics relating to a field of study are not always in synchrony, which can introduce unfortunate — and unacceptable — restrictions.

The Corrigendum relates to a paper that describes making pluripotent germline stem cells from human testicular tissue (S. Conrad *et al. Nature* **456,** 344–349; 2008). As with all materials or data reported in a Nature research paper, these cells must be shared with the research community on publication. However, as the Corrigendum details, none of the pluripotent cell lines is currently available because of restrictions in the donors' consent forms, which limit distribution and promise that the cells will be destroyed after a certain time in culture. The authors have since contacted the donors and received permission to develop a few of the lines for distribution. They are also making new lines for distribution from other donors.

Such failures to distribute cell lines are incompatible with Nature journal policies and with the efficient progression of scientific knowledge. The Corrigendum alerts investigators to this situation and the steps being taken to rectify it. This particular case illustrates how, even when clinicians, researchers and their local ethics board follow internal procedures that promote both donor safety and medical research, serious problems can arise.

The community was not that surprised by this situation — six of seven researchers contacted by *Nature* thought this could happen again. Researchers developing cell lines must investigate the restrictions associated with the human tissue they are using, particularly if someone else collected the samples, if the samples come from multiple clinical sources or if they come from several legal jurisdictions. If a scientist needs to create cell lines that might be used for as-yet-unforeseen purposes, only tissue with no restrictions should be used. An article published earlier this year helpfully suggests that scientists obtaining human tissue could alleviate most of the issues from the start by mentioning in the consent form some common procedures (such as sharing cell lines with other investigators), and by including a request to contact donors again if their research takes unanticipated directions (K. Aalto-Setälä *et al. PLoS Biol.* **7,** e1000042; 2009).

Journals can remind authors in their policy guidelines that authors of submissions that involve consent forms must make editors aware of any limits that result from those forms. The Nature journals will be revising their policies to make this clearer.

Most importantly, patients, researchers, clinicians, and review and ethics boards worldwide need to agree on conventions that are acceptable to most parties under most circumstances. Internationally standardized consent forms for the donation of human tissue should cover new uses, genomic comparisons, patents and product development, and should discourage limiting access or lifespan.

Ethics and review boards are set up to protect individuals, but can also go much further to promote research. No one can deny that donors need to understand the risks and benefits of a procedure, trial or donation. However, it seems most ethically responsible, given the value of research, for the boards to explain the consequences that restricted access and time limits can have on the value of a donor's tissue. ∎

A question of control

Scientists must address the ethics of using neuroactive compounds to quash domestic crises.

A number of countries are investigating the use of neuroactive compounds as a nonlethal way to deal with riots and other domestic crises. The idea is to stun people temporarily, or otherwise change their behaviour, to help the authorities exert control (see page 950).

Russian special forces put that idea into practice in October 2002, when they sprayed a mixture of incapacitating agents into a Moscow theatre in an effort to free some 700 theatre-goers held captive by Chechen rebels. The exact nature of the mixture remains secret; Russian authorities disclosed only that it included a component similar to the opiate fentanyl. But it obviously had a narrow therapeutic window: about 130 hostages died as a result of inhaling the gas.

This episode underscores the ethical conundrum — would the rebels have killed all the hostages? — that makes an outright ban on the military use of incapacitating agents politically unrealistic. Instead, an acrimonious argument over the control of nonlethal weapons is now under way among the states that have signed the Chemical Weapons Convention — which does not cover nonlethal uses for domestic riot control and the like — as well as the Biological and Toxin Weapons Convention, which states that biological agents may be used only for "prophylactic, protective or other peaceful purposes". Unfortunately, the various sides cannot even agree on how to define the exclusions in the treaties.

During this impasse, the wider community of life scientists should actively discuss the effectiveness and safety of potential incapacitating agents. In particular, academics and non-governmental organizations involved in the debate should agree a list of compounds likely to be considered for use by military agencies, and publish it on the Internet. Scientists could then submit comments to aid its annotation.

Just listing potential agents does not necessarily imply an endorsement of their use. But by providing an accessible forum where scientists can directly engage on the issue of nonlethal weapons, it could help inform the political debate — and might prevent disasters of the sort seen in Moscow. ∎

RESEARCH HIGHLIGHTS

Pearly pedigree

Science doi:10.1126/science.1173793 (2009)

The iridescent sheen of nacre — the mother of pearl that coats the inner shell of molluscs — arises from the light-scattering properties of its finely layered structure. This comprises crystals of aragonite — a form of calcium carbonate — aligned within organic molecules. Hiromichi Nagasawa and his colleagues at the University of Tokyo have now discovered two proteins key to the formation of this material.

Studying the shell of the Japanese pearl oyster (*Pinctada fucata*), the researchers found that one protein, Pif 80, binds specifically to the aragonite crystals. When calcium carbonate was crystallized with organic material lacking Pif 80 and another protein, Pif 97, aragonite crystals grew randomly, with no iridescent effect.

The team thinks that the two proteins form a complex that helps to guide the formation of aragonite crystals, which are then assembled within sheets of organic material.
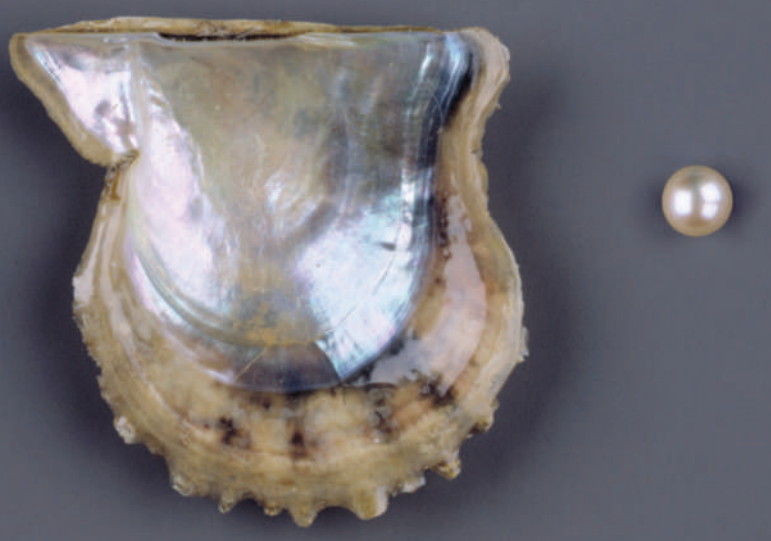
D. FINNIN/AMNH/MASTOLONI PEARL COLLECTION

ANIMAL BEHAVIOUR
Ties that bind

PLoS ONE 4, e6573 (2009)

Rescue attempts by ants were first documented more than 100 years ago. Now Elise Nowbahari at the University of Paris–North and her colleagues report that such efforts are undertaken only for struggling colony mates.

The researchers used a nylon thread to tie a *Cataglyphis cursor* ant to filter paper, and partially buried the ant under sand. They found that only when active colony-mates were tied to the paper did other ants come to the rescue; chilled, motionless ants, or those of different colonies or species, elicited no aid.

The rescue ants dug sand until they uncovered the snare, which they then bit at. The sophistication of this snare-biting, the researchers add, far outdoes previously reported rescue activities such as sand-digging and leg-pulling, assumed to be responses to chemical alarm calls from any ant.

PHYSICS
Close heat

Nano Lett. doi: 10.1021/nl901208v (2009)

The flow of heat between two surfaces held sufficiently close and separated by a vacuum can, in theory, exceed the limits laid down by Max Planck's law of blackbody radiation.

Gang Chen and Sheng Shen at the Massachusetts Institute of Technology in Cambridge and Arvind Narayanaswamy, now at Columbia University in New York, have observed this increased heat transfer in an experimental feat. They measured the electromagnetic radiation passing between a glass sphere and a flat glass, silicon or gold plate held as close as 30 nanometres.

The increase in radiation is due to electromagnetic waves that, unusually, do not radiate outwards but instead ripple over the surface of polar materials such as glass, and are formed by a combination of photons and lattice vibrations.
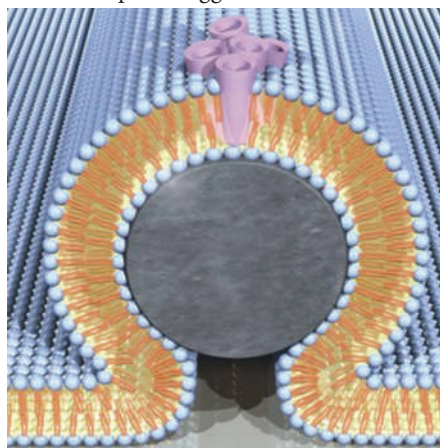
BIOELECTRONICS
It's electrifying

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0904850106 (2009)

Ions and membranes — common biological signal transducers — have been combined with electronic circuitry.

Aleksandr Noy of the Lawrence Livermore National Laboratory in California and his team insulated 20–40-nanometre-wide silicon nanowire transistors by coating them with lipid membranes (pictured below), then added proteins that formed channels through the membranes. A flow of ions through the membrane pores triggered an electrical response in the nanowire. Conversely, by applying an electric field using the wire, the researchers could open and close the membrane's protein channels.

J. MARTINEZ, N. MISRA, A. NOY AND S. DOUGHERTY (LLNL)

POPULATION ECOLOGY
Evolution to the rescue

Ecol. Lett. 12, 942–948 (2009)

Environmental degradation threatens to wipe out countless species unless natural selection can help them to adapt fast enough to fend off ultimate doom — a process known as 'evolutionary rescue'.

Theory predicts that large populations are necessary to keep extinction at bay, but the mathematical models had never been tested. Graham Bell and Andrew Gonzalez of McGill University in Montreal, Canada, exposed hundreds of populations of baker's yeast (*Saccharomyces cerevisiae*) to deadly salt concentrations. Consistent with prediction, the microbial populations crashed before rapidly bouncing back, and recovery occurred only when the population had started out above a threshold size.

NEUROSCIENCE
Categorically hard-wired

Neuron 63, 397–405 (2009)

The human brain uses different regions of the visual cortex to process recognition of living and inanimate objects. Researchers have now shown that this doesn't depend entirely on the ability to see the objects: it occurs even in people who have been blind since birth.

Alfonso Caramazza at Harvard University, Bradford Mahon at the University of Rochester in New York and their colleagues read out lists

of six animals to congenitally blind individuals and sighted people, all with closed eyes, asking them to compare the animals' sizes. Using functional brain imaging, the researchers showed that the lateral occipital cortex became active in both groups. But when lists of non-living objects, such as tools, were read out, a different area of the visual cortex — the medial fusiform gyrus — was activated.

Evolution may have selected for hard-wiring that separates neural categories for animals — towards which humans have important emotional responses — from those for non-living things, the authors say.

CANCER
From the source

Nature Med. **15**, 907–913 (2009)

Tumours in a particularly aggressive form of breast cancer may originate from cells lining the mammary duct — not from mammary stem cells as previously thought.

Jane Visvader and Geoffrey Lindeman at the Walter and Eliza Hall Institute of Medical Research in Victoria, Australia, and their colleagues compared the breast tissue of healthy women with and without mutations in the gene *BRCA1*, which significantly increase the risk of breast cancer.

Tissue with mutated *BRCA1* harboured a larger population of luminal progenitor cells, and these were able to proliferate without the growth factor necessary for the division of other cell types, including mammary stem cells. Uncontrolled growth is a feature of cancer cells. The authors also found that normal luminal progenitors, precancerous tissue from *BRCA1* carriers and basal breast tumours had similar gene-expression profiles.

BIOPHYSICS
Protein friction

Science **325**, 870–873 (2009)

Motor proteins that carry cargo around the cell by 'walking' along microtubule filaments are slowed by the friction caused by the rupturing of chemical bonds as they go.

Jonathon Howard of the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany, Erik Schäffer of the Technical University of Dresden and their colleagues measured this friction *in vitro*. Using optical tweezers, they dragged a microsphere coated in kinesin-8 — a motor protein from yeast — over an immobilized microtubule. They calculate that, at top speed, friction between a kinesin molecule and its filament dissipates about half of the chemical energy that the cell usually supplies to drive the motor along its track.

IMMUNOLOGY
Helping the helpers

Immunity doi:10.1016/j.immuni.2009.08.001 (2009)

In the immune system, 'helper' T cells secrete signalling molecules called cytokines to activate other immune cells. A subset of helper cells, $CD4^+$ T_h17 cells, can trigger autoimmune diseases such as multiple sclerosis when they release the proinflammatory cytokine IL-17.

Kingston Mills and his colleagues at Trinity College in Dublin have now implicated another type of T cell, the $\gamma\delta$ T cell, in autoimmunity. The group showed that $\gamma\delta$ T cells are an innate source of IL-17 and other signalling molecules, which they produce when activated by particular cytokines *in vitro* and in live mice.

The authors found IL-17-secreting $\gamma\delta$ T cells in the brains of mice with an autoimmune disease — and that transplanting these alongside $CD4^+$ cells conferred disease on other mice. *In vitro*, activated $\gamma\delta$ T cells amplified the production of IL-17 by $CD4^+$ T_h17 cells.
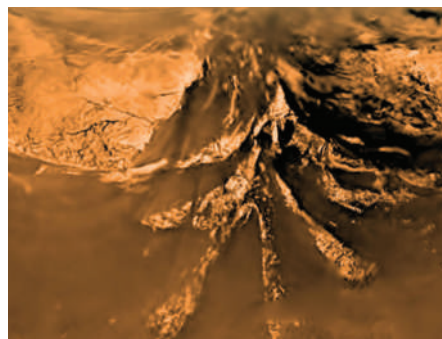
ESA/NASA/UNIV. ARIZONA

PLANETARY SCIENCE
What an atmosphere

Icarus doi:10.1016/j.icarus.2009.07.040 (2009)

The atmospheric shroud of Saturn's large moon Titan includes methane — lending it its orange hue (picured above) and an element of mystery, because sunlight breaks the gas down relatively quickly. If Titan's store of methane were limited to its lakes and rivers, it would have lost the gas from its atmosphere long ago. So where is it coming from?

Olivier Mousis at the University of Arizona in Tucson and his colleagues rule out one idea: that a reaction between subsurface water and rock is continually replenishing the methane. The ratio of deuterium and hydrogen in the gas is not consistent with a geological origin, they find.

Instead, their analysis suggests that Titan scooped up all of its methane 4.5 billion years ago when it was formed — and could have as much as 1,300 times the amount seen in its atmosphere still trapped in its interior.

JOURNAL CLUB

Heather Stoll
Department of Geology,
University of Oviedo, Spain

A biogeochemist sees the value of diversity in a changing ocean.

Ocean acidification in response to excess carbon dioxide in the atmosphere could become a problem for marine organisms, especially those that make skeletons or shells out of calcium carbonate. Corals and clams are at risk, as are the coccolithophorids — microscopic algae that are, by volume, the most important shell producers.

These algae have been the guinea pigs in a series of lab studies measuring their response to acidified seawater. But I worry about whether these studies give us an accurate picture of the future. They typically start with clones — descendants of a single cell — grown in acidified conditions for only a few weeks. This set-up precludes the kind of natural selection and adaptation that might occur over decades and centuries in the ocean.

To cloud the waters further, different labs often obtain conflicting results on the same species, a situation some attribute to subtle differences in methods. Fortunately, a recent study by Gerald Langer of the Autonomous University of Barcelona in Spain and his colleagues provides a more satisfying and ultimately more optimistic explanation (G. Langer *et al. Biogeosci. Discuss.* **6**, 4361–4383; 2009). These reserachers grew four different strains of a calcifying algae, *Emiliania huxleyi*, at different seawater pH levels, and showed that the response to acidification varies significantly among the strains. They argue convincingly that these diverse responses have a genetic basis.

Identifying diverse responses among strains of a species puts us one step closer to capturing the true potential of adaptation in this group of organisms. It would be naive to assume that this puts coccolithophorids out of harm's way. However, diversity is good insurance in a changing ocean. Moreover, I am hopeful that scientific experiments are starting to take that into account.

Discuss this paper at **http://blogs. nature.com/nature/journalclub**

NEWS

Paying to save the rainforests

In Brazil, details are emerging for plans to stop deforestation. Can it serve as a model for other nations?

Along the Trans-Amazonian Highway in the Brazilian state of Pará, many landowners try to boost their income by clearing a hectare or two each year for farms or cattle grazing. This year, however, may be different: if all goes to plan, around 350 families will receive payments to put rainforest preservation first.

If approved by the Brazilian Development Bank within the coming weeks, the project would be one of the first to stem from the Amazon Fund, a major initiative created by Brazil last year to attract international aid. It is the largest forest-conservation initiative in the world, and the only national programme that could demonstrate how tropical-forest protection might be folded into the global-warming treaty that international leaders hope to sign in Copenhagen in December.

"The Amazon Fund could be interpreted as the fundamental test case for the rest of the world," says Paulo Moutinho, who heads research at the Amazon Institute for Environmental Research (IPAM) in Brasília. "The international community is watching Brazil and how we will deal with this experiment."

With a total price tag of about US$17 million, the Pará project is just one example of how the Amazon Fund could distribute its money. Landowners who sign up to preserve their forest would begin receiving monthly cheques, starting at around $16 and increasing to $350 in the tenth and final year. Other investments would help to modernize local agriculture in an effort to increase income from land that is already cleared, so that landowners don't need to begin cutting down trees again when the payments stop. Moutinho says the programme could be scaled

Cash incentives could be part of a multi-pronged approach to stopping slash-and-burn deforestation.

up to 10,000 families in the Pará region alone.

All told, the project would reduce the otherwise expected greenhouse-gas emissions by 3.1 million tonnes of carbon dioxide — equivalent to taking more than half a million vehicles off the road for one year — at a cost of just more than $5 per tonne. That is 75% less than the going price on the European carbon market. Backed by a satellite monitoring system and an increasingly focused enforcement programme, Brazil thus has an opportunity to show whether this way of reducing emissions from deforestation and degradation (REDD) works.

"I would call the Amazon Fund the biggest experiment in tropical conservation history," says Dan Nepstad, a researcher at the Woods Hole Research Center in Massachusetts. "If it works, REDD will survive. If it fails, there's a chance REDD will fail."

The Amazon River basin covers some 7 million square kilometres and nearly half of Brazil. By some estimates 15% of the basin has been cleared in recent decades. Worldwide, deforestation accounts for as much as 20% of greenhouse-gas emissions, and up to 70% of Brazil's emissions. Climate negotiators in the

**"The Amazon Fund is the biggest experiment in tropical conservation history."**

United Nations talks are looking at various ways to link international carbon markets to forest conservation, but Brazil has long opposed the idea of allowing US or European companies to offset their emissions by paying for forest conservation projects in the tropics.

The Amazon Fund was designed as an alternative, allowing Brazil to deploy direct international aid as part of a comprehensive national strategy. Last year Brazil pledged to reduce deforestation by 70% by the end of 2017; the government has since extended that commitment to 80% by 2020.

Achieving those goals won't be easy, given poverty levels, enforcement difficulties and ongoing questions about who holds title to what land. Even within the federal government, policies promoting agricultural growth are often at loggerheads with those intended to protect rainforest. Deforestation rates fell for three years after peaking in 2004 (see graphic), but then increased in the 2008 season when prices spiked for commodities such as soya and beef. Deforestation rates seem to have dropped again in the most recent season; experts credit better enforcement and new policies but also the economic crisis,
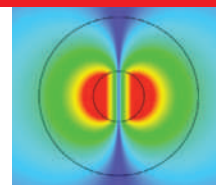

DEFORESTATION IN THE BRAZILIAN AMAZON

Deforestation rate (thousands of square kilometres)
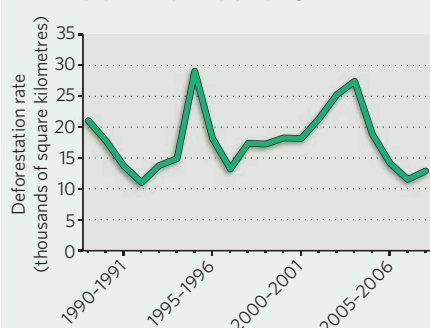
Source: Instituto de Pesquisa Ambiental da Amazônia

G. GILABERT/CORBIS SABA

which cut demand for many commodities.

The Amazon Fund got off the ground with a pledge from Norway, which committed up to $1 billion until 2015. Brazil will receive around $114 million this year, but must continue reducing emissions in order to receive future payments. Climate negotiators are increasingly focusing on national baselines such as this, instead of on particular projects that might save one patch of forest while pushing loggers, developers and landowners down the road to another patch.

Getting REDD right in Brazil and beyond is "totally possible and essential", says Lars Løvold, director of the Rainforest Foundation Norway in Oslo, which, along with Friends of the Earth Norway, proposed to the Norwegian government that it invest in a big forest conservation initiative. "But you need some projects to show that it works."

Eyes in the sky

In the coming weeks, the Brazilian Development Bank, which manages the Amazon Fund, is expected to announce the first such project awards. Several dozen applications have been submitted, ranging from community initiatives like the project in Pará to land registry programmes and a proposal from Brazil's National Institute for Space Research for a new satellite to monitor deforestation.

Within Brazil, the money coming from abroad has whetted local appetites for more. And in June, the nine governors of the Brazilian states in the Amazon region penned a letter to Brazilian President Luiz Inácio Lula da Silva urging the country to reconsider its opposition to directly tapping carbon markets for forest conservation. The governors called the Copenhagen talks "a golden opportunity", suggesting that carbon markets could surpass $2 trillion annually by 2020 and $15 trillion in 2050.

Paulo Adario, Amazon campaign director for the Brazilian arm of environmental organization Greenpeace, is wary of governors opening their states directly to international investments; such a deal, he says, could undermine the idea of a national baseline, without which there is no way to protect the forest as a whole. "The federal government needs to have a national vision about the problems and the solutions for the country," he says, "and then performance will be evaluated against results."

The official deforestation data for the 2009 season, which ended in July, will be available in December. Preliminary results suggest that total deforestation will hit a two-decade low of less than 10,000 square kilometres — low enough to secure another payment from Norway in 2010. ∎
Jeff Tollefson

Nanoparticle safety in doubt

Claims that seven Chinese factory workers developed severe lung damage from inhaling nanoparticles are stoking the debate over the environmental-health effects of nanotechnology.

A paper published in the *European Respiratory Journal* claims to be the first to document cases of ill health caused by nanoparticles in humans (Y. Song, X. Li and X. Du *Eur. Respir. J.* 34, 559–567; 2009). Other experts are sceptical as to whether nanoparticles are actually to blame, but the paper has triggered lively discussions.

"The study raises the bar for doing appropriate research as fast as possible to find out where the dangers might lie when working with nanomaterials," says Andrew Maynard, a nanotechnology expert at the Woodrow Wilson International Center for Scholars in Washington DC.

The study describes seven women, aged 18–47 years, who worked in an unidentified printing factory in China; two of them later died. They all had pleural granulomas — ball-like collections of immune cells in the lining of the lung that form when the immune system is unable to remove a foreign body. They also had excessive, discoloured fluid in the lung lining. Particles around 30 nanometres in diameter were found in lung fluid and tissue.

The study says that the symptoms were caused by inhaling fumes produced when the workers heated polystyrene boards to 75–100 °C. The boards had previously been sprayed with a 'paste material' made from a plastic identified as a polyacrylate ester.

The workroom, of around 70 square metres, had one door and no windows. The ventilation unit had broken down five months before symptoms started to manifest, and the door had been kept closed to keep the room warm. The workers wore cotton gauze masks only on an "occasional basis".

Electron microscopy found nanoparticles around 30 nanometres in diameter in the paste and in dust particles that had collected at the inlet of the broken ventilation unit. Lead author Yuguo Song, a clinical toxicologist at Beijing Chaoyang Hospital, says "it is obvious the disease is not due to microparticles or vapours, because the pulmonary epithelial cells are

Could nanoparticles cause some lung damage?

full of nanoparticles".

Maynard says the symptoms seen in the patients are "similar" to those seen in animals exposed to nanoparticles. He adds that damage to the areas surrounding the lungs suggests that larger particles are not to blame, as these tend to be constrained within the lungs. But because the study does not identify what nanoparticles were involved or their concentration, he says, "we can't say what the link is or if there are other exacerbating circumstances".

Ken Donaldson, a respiratory toxicologist at the University of Edinburgh, UK, doubts that nanoparticles are to blame. He says the symptoms are more typical of chemical exposure. "I don't doubt that nanoparticles were present, but that does not mean they were the main arbiters," he says.

Donaldson says that the plastic material the patients worked with is the more likely culprit — as it would have been highly toxic at the levels they were probably exposed to given the size of the room they worked in and its lack of ventilation.

Anthony Seaton, an emeritus professor in environmental and occupational medicine at the University of Aberdeen, UK, agrees that the study does not pin down nanoparticles as the cause of the ill health. Rather than an insight into the toxicology of nanoparticles, he says, the study is an example of a "total failure in health and safety procedures". ∎
Natasha Gilbert

> **"We can't say what the link is or if there are other exacerbating circumstances."**

Japan election sparks science pledges

Both parties make vague promises about research, but differ in their climate targets.

Japan's upcoming elections, on 30 August, could see the party that has ruled almost continuously for more than half a century lose its grip on power. As *Nature* went to press, polls suggested that the Democratic Party of Japan (DPJ) would take a majority in the lower house of parliament, giving it the right to appoint a prime minister to replace Taro Aso of the Liberal Democratic Party (LDP).

Some researchers have expressed concern that the DPJ, with its emphasis on the responsible use of taxes, might cut basic-research budgets. Last week at a press conference, DPJ president Yukio Hatoyama tried to dispel such rumours. "Japan should be a leading country in basic science," he said in response to a question from *Nature*. "If anything, I think overall scientific funding should be expanded."

The DPJ has been throwing out similar promises to voters concerned about Japan's struggling economy. For example, the party's manifesto includes a new annual allowance of ¥312,000 (US$3,300) for every child until they finish junior high school. The party says that it can pay for this and other big programmes by reducing spending elsewhere, such as on road and dam projects, and shifting decision-making power towards elected officials and away from government bureaucrats.

But Japan last week officially pulled out of recession, and many scientists have been happy with the way bureaucrats have fought to maintain or even increase the science budget during slow economic growth (see chart) when most other sectors of government were seeing budgets cut. Major projects include the Japan Proton Accelerator Research Complex in Tokaimura and the X-ray free-electron laser scheduled to be completed next year at the

Yukio Hatoyama, president of the opposition DPJ.

SPring-8 synchrotron in western Japan.

Biophysicist Akiyoshi Wada, former director of RIKEN's Genomic Sciences Center in Yokohama, speculates that the DPJ's declared drive to get rid of waste might affect research. "They say they want to stop listening to the bureaucrats, but how much will they be able to do without?" Wada asks. "Large-scale 'challenging' projects might get drastically cut." Indeed, the party's manifesto promises to re-evaluate the "independent administrative organizations" that include many of Japan's largest research bodies, among them RIKEN.

In response to a *Nature* questionnaire, Hatoyama's office pledged support for big projects: "Japan is already a top runner in some fields that require huge budgets, like particle physics," it said, "and we will actively aim to maintain that position by building world-class research bases and deepening research ties with Europe, the United States and Asian countries." The statement also said that the DPJ would take (unspecified) measures to increase the science budget. It noted that although Japan overall invests a high percentage of its gross domestic product (GDP) in research — 3.67% in 2007 — the percentage of government investment is low compared with investment by industry. "Cultivating researchers needs to be part of our central policies," the statement said.

Hatoyama received a doctorate in engineering from Stanford University in California and was a professor at the Tokyo Institute of Technology. He plays down his scientific background,

saying "what matters for politicians is having principles and convictions, and a broad perspective". Naoto Kan, the former DPJ head and one of its two acting presidents, got a doctorate in applied physics before becoming a patent lawyer; some say he might bring more attention to issues that have plagued Japanese patent law, such as the need for broader patents to compete with the United States and streamlining the lengthy patent application procedure.

Hatoyama says that the DPJ will create a 'science and technology strategy office' to replace the Council for Science and Technology Policy (CSTP), the country's highest science-policy body. Like the CSTP, the new office would coordinate science-related budget requests, but it would also, in ways unspecified, take a more active role in "promoting basic and applied sciences in a unified manner".

If the DPJ does create a top-level scientific body, this might find itself in the thick of an ongoing debate over the role of big science projects. Tadamitsu Kishimoto, an immunologist at Osaka University and a former CSTP member, says money has been wasted on big biology projects, and he complains that the ¥270 billion earmarked for 30 new projects by the CSTP could be more of the same. "Right now young PIs don't get enough to work with," he says, "and using that money to give ¥10-million grants to many young researchers would produce a lot more good science."

The LDP platform acknowledges the many Nobel prizes garnered by Japanese scientists, and vows to "educate and send out into the world more researchers than ever before", and to create 30 "world-class research facilities".

On climate change, the DPJ calls for a cut in greenhouse-gas emissions to more than 25% below 1990 levels by 2020. The LDP's proposal of an 8% target has been criticized by developing countries and environmental groups as not going far enough. But Japan has consistently missed targets for its original Kyoto Protocol pledge to cut emissions to 6% below 1990 levels during 2008–12, and many in the business community in particular are sceptical about the DPJ's emissions targets.

"If we follow the DPJ's targets, we will end up with a 3.2% drop in the GDP" based on government estimates, LDP secretary-general Hiroyuki Hosoda said last week. "The party that presents the more dramatic figure is more attractive, but it is not feasible," he said. ∎

David Cyranoski

JAPAN'S SCIENCE & TECHNOLOGY BUDGET

SOURCE: MEXT
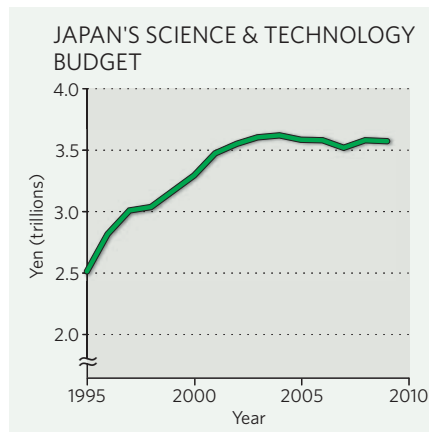
CHILD DNA RAISES CONSENT QUESTIONS
Bioethicists argue for stricter biobanking rules.
www.nature.com/news

CORBIS

Collins sets out his vision for the NIH

BETHESDA, MARYLAND
On his first day in the job, the new director of the US National Institutes of Health (NIH) laid out a five-point road map for the agency — which includes focusing greater attention on translational research, neglected diseases and health-care reform. But Francis Collins's top priority will be tackling budget constraints after the $10.4-billion boost from the economic stimulus package runs out in 2010.

"I don't want you to think that all it's going to take is a few speeches or maybe a little playing the guitar for this to be successful," he said at a 17 August town-hall meeting at NIH headquarters in Bethesda.

For example, he estimated that only around 3% of applicants for the stimulus-backed 'challenge grants' will be funded, although some have predicted success rates of below 1%. Collins has raised red flags about losing a generation of young scientists if the NIH budget drops or flatlines. The budget for fiscal year 2010 has not yet been finalized, but the administration of President Barack Obama has requested

a 1.4% increase over the 2009 budget of $30.6 billion.

Collins, who headed the National Human Genome Research Institute in Bethesda from 1993 to 2008, made the case that investment in biomedical research creates jobs and offers quick economic returns. Looking ahead, he said that the agency should devote more money to "five areas of special opportunity". First, to applying high-throughput technologies in genomics and nanotechnology to discover the genetic bases of diseases including cancer, autism, diabetes and neurodegenerative disorders. Second, to developing diagnostics, preventative strategies and therapeutic tools through more public–private partnerships. Third, to reining in the costs of health care with comparative-effectiveness research and personalized medicine. Fourth, to expanding research into diseases affecting the developing world. Finally, to increasing budgets and investing in training and peer review to achieve a predictable funding trajectory for the research community.

"He presented a vision that was very much

tailored to the times — scientifically and in terms of our public-health challenges, of where the nation is now, and where the mission of the NIH lies ahead," says Raynard Kington, who has served as acting director since last October and will return to his role of deputy director.

Collins stressed that he plans to devote all his energy to his new position, although he will maintain an active research lab. His latest book (on personalized medicine) is slated for a January 2010 release, he said, and he has resigned from the Washington DC-based BioLogos Foundation, an organization he founded earlier this year aimed at bridging the gap between science and religion. "I don't want anyone at NIH or outside of NIH to think that I have a religious agenda in coming to this role," said Collins, a devout Christian. "I will not. I'm here as your scientific director and I will keep that focus."

However, Collins's wife, genetic counsellor Diane Baker, will stay on the BioLogos board and will probably have a significant advisory role at the NIH, he said. ∎

Elie Dolgin

Environmental concerns delay seismic testing

An environmental lawsuit is threatening the departure of a long-planned, US$4.7-million research cruise to image sea-floor structures off the coast of western Canada.

The RV *Marcus Langseth*, a vessel operated by Columbia University in New York for the US National Science Foundation, had acquired all its permits to depart on 21 August for the Endeavour hydrothermal vents, 250 kilometres southwest of Vancouver Island. But on 10 August, the Canadian activist legal group Ecojustice, in Vancouver, British Columbia, sued the university, the Canadian department of fisheries and oceans and the minister of foreign affairs, alleging among other things that proper procedures were not followed in assessing how the seismic air bursts set off during the cruise would affect marine life.

Cruise co-leader Douglas Toomey, a geophysicist at the

US research vessel *Marcus Langseth* was scheduled to work in Canadian waters.

University of Oregon in Eugene, says the ship was specifically scheduled and routed to avoid whales, and that marine-mammal observers would be on hand when airguns were fired.

As *Nature* went to press, a hearing on the lawsuit was being scheduled for this week.

The cruise plans to sink 64 portable seismometers near

the Endeavour vent field, fire airguns into the water for 10 days, then retrieve the data-laden seismometers that pop to the surface. Toomey hopes that the analysis will answer important questions about the flow of Earth's mantle, as well as about earthquakes in the region.

Ecojustice officials didn't respond to interview requests, but

a statement on its website attributed to group lawyer Lara Tessaro says that the Canadian government should "refuse to sanction the harassment of endangered whales". The lawsuit was filed on behalf of the Living Oceans Society and the Canadian Parks and Wilderness Society, and argues that the minister of foreign affairs should not grant the vessel clearance.

The use of airguns for seismic studies in western Canada has stirred environmental anger before. In 2008, a joint US–Canadian study of granite structures underlying British Columbia was halted (see *Nature* 451, 3; 2008). Since then, the region has seen growing environmental activism, including a recent sabotage (see *Nature* doi:10.1038/news.2009.715; 2009) of an onshore explosive charge that was part of a seismic test. ∎

Rex Dalton

LAMONT–DOHERTY EARTH OBSERVATORY

SPECIAL REPORT

Cardiovascular disease gets personal

Gene-association studies hint at better ways of treating the leading cause of death, but capitalizing on them is proving to be a slow and difficult process. **Erika Check Hayden** reports.

As personalized cancer treatment edges into the clinic, doctors and scientists are hoping that cardiovascular disease — the world's top killer — will be next to benefit from genomics.

An avalanche of studies has linked genetic variants to various cardiovascular conditions and to patients' responses to commonly prescribed drugs. First up could be genetic guidance for the anti-clotting agents warfarin and clopidogrel, followed by testing for genetic variants responsible for conditions such as atrial fibrillation, a heart-rhythm abnormality that is a leading cause of stroke.

Doctors caution that there is a long way to go before the hints raised by gene-association studies translate into solid evidence that genetic variants can improve clinical practice. "I would hope that cardiovascular disease would be one of the next leading areas of personalized medicine, because it has such an enormous impact on public health," says Christopher Granger, a cardiologist at Duke University Medical Center in Durham, North Carolina. But "we're at a primitive stage right now", he says.

Warfarin exemplifies some of the promise and pitfalls of personalized medicine. The drug is commonly used to prevent clotting in patients who have atrial fibrillation and other conditions, yet the dose needed varies from patient to patient; if it's not precisely right, it can trigger fatal haemorrhaging. In 2007, after studies found that variants in two genes, *CYP2C9* and *VKORC1*, account for up to half of the reason why patient response differed, the US Food and Drug Administration (FDA) changed the labelling to suggest that doctors consider using genetic tests to guide dosing.

Yet studies have failed to show that such tests help improve patient outcomes. In 2007, for instance, a trial of 206 people reported that using information about a patient's genetic variants to guide their dosing regimens didn't lessen the risk that patients on warfarin would develop unsafe levels of clotting proteins[1]. And this January, another group reported that genetic testing wouldn't save money if done in all patients prescribed the drug, partly because it still costs hundreds of dollars to determine the genetic variants of each patient[2].

The US National Heart, Lung and Blood Institute in Bethesda, Maryland, is now sponsoring a larger clinical trial to test the usefulness of

Cardiovascular conditions are the leading cause of death worldwide.

genetically guided warfarin dosing. But as Eric Topol, director of the Scripps Translational Science Institute in La Jolla, California, says: "Warfarin was kind of the poster adult for pharmacogenomics, but it's really lost favour."

He and other doctors now see greater potential instead for clopidogrel, marketed by Bristol-Myers Squibb of New York and Sanofi-Aventis of Paris as Plavix. Clopidogrel is given to fight clotting, including in patients who have already had a stroke or heart attack. The drug, second in the world in global sales, is converted in the body into an active form that inhibits the pro-clotting protein P2Y12. But it, too, can cause haemorrhages.

In December, three groups reported that variations in the *CYP2C19* gene were associated with an increased risk of cardiovascular events in patients on clopidogrel[3–5], and a poor ability to convert clopidogrel into its active form[4].

And last month, the FDA approved a new drug, prasugrel, which is a more potent inhibitor of P2Y12 but also carries a higher risk of bleeding. It is conceivable that patients who have the genetic variants associated with poor response to clopidogrel might instead be

treated with prasugrel, says Matthew Price, an interventional cardiologist at Scripps Clinic/ Green Hospital in La Jolla.

Misunderstood

Predicting the overall risk of cardiovascular diseases is proving even more complicated than treating them with genetically targeted drugs. Many genome-wide association studies have been done, but few have uncovered variants that, on their own, boost the risk of cardiovascular disease very much. And taken together, the variants discovered so far still don't explain most of the genetic risk of various diseases.

For instance, last January, US researchers aimed to improve risk prediction of cardiovascular disease by adding information about a genetic variant associated with coronary artery disease and diabetes to other risk factors, such as smoking, cholesterol levels and family history of heart attack. The variant, found on chromosome 9, had formed the basis of a genetic test sold by deCODE Genetics in Reykjavik, Iceland. The team found that genotyping the variant did not improve the ability to predict

whether the 22,129 women in the study would develop heart disease[6]. "Our study didn't show very much change [in risk-prediction ability], especially over the risk score that had family history in it already," says Nina Paynter, team leader and an epidemiologist at Brigham and Women's. "I was a little bit disappointed."

Since Paynter's study began, however, many more genetic associations with cardiac risk have been reported, and companies such as deCode, Navigenics of Foster City, California, and 23andMe of Mountain View, California, now sell tests that purport to assess heart-disease risk using combinations of these variants. Paynter's group is evaluating multi-variant genomic tests, as is the independent Evaluation of Genomic Applications in Practice and Prevention initiative set up by the US Centers for Disease Control and Prevention in 2004, which is expected to issue recommendations on their use this autumn.

Other groups have already examined genomic tests for diabetes, which greatly increases the risk of heart disease and stroke. Last year, for instance, three groups published studies examining whether a number of variants associated with diabetes could predict a person's risk of developing this disease[7-9]. All found that the variants added little to the predictive value of known diabetes risk factors, such as obesity, smoking and family history.

In addition, the way these genomics tests are reported can be confusing to consumers. Companies update consumers' risk profiles as new variants are discovered, but because each new variant changes a person's risk so little, variants added to a risk profile can cancel out previous ones.

A team led by Cecile Janssens of Erasmus Medical College in Rotterdam, the Netherlands, showed this by studying the same diabetes-associated variants analysed in the 2008 risk-prediction studies[10]. When genotypes of 17 of these variants were added to an existing risk profile based on variants of *TCF7L2* — a gene whose variants confer a substantial increase in risk of common forms of diabetes — 34% of the patients' risk profiles changed, for example, from high to low or low to high. When data about patients' age, sex and body mass index were added to the profiles, 29% changed risk categories, and 11% of the participants reverted to their initial risk category.

Patients hoping to use their genotyping results to motivate healthy lifestyle changes might thus be confused when their disease risk changes multiple times without any action on their part, Janssens says: "Our studies show that

"I would hope that cardiovascular disease would be one of the next leading areas of personalized medicine."

these products are not ready for prime time."

Cardiologists hope that will change, and see some promise on the horizon in specific cardiovascular diseases.

Last month, for instance, two research teams published studies that linked variants in the *ZFHX3* gene to atrial fibrillation[11,12]. Two years ago one of the same groups published variants adjacent to a separate gene, *PITX2*, that almost doubles the risk of atrial fibrillation[13]. Drugs and monitoring can be used to treat the condition, and might help prevent the roughly one-third of strokes that have no known cause. So atrial fibrillation could serve as an early example of genomic risk prediction, Topol says.

"To be able to zoom in on the probable cause of a stroke by genomics, and then institute a much more intensive heart-rhythm monitoring programme, would be a whole new path that we didn't have months or even a year ago," he says.

But even if the tests are proven useful, they may still have a limited impact on patient care — at least at first. Granger says that doctors already have various risk-prediction tools that work, but don't use them effectively. For instance, patients with higher levels of the protein complex troponin, an indicator of heart-muscle damage, do better on certain treatment strategies. But, Granger says, they are no more likely to get those drugs for various reasons, including that family doctors are not as familiar with the cardiac literature. "Part of our challenge is that we've already got some information that could help us better customize medicine to patient risk, and we tend not to be doing that in practice."

Remedying that problem will require more physician education — and some knockout examples of genetic profiling aiding medicine, as seems to be happening in cancer. "Maybe we need a couple of major success stories of the benefit of using genetic variants for treatment of disease, and I think we're getting some of those," Granger says. Cancer drugs may be blazing the trail for personalized medicine, but cardiovascular drugs may not be far behind. ■

1. Anderson, J. L. *et al. Circulation* **116,** 2563–2570 (2007).
2. Eckman, M. H., Rosand, J., Greenberg, S. M. & Gage, B. F. *Ann. Intern. Med.* **150,** 73–83 (2009).
3. Collet, J.-P. *et al. Lancet* **373,** 309–317 (2009).
4. Mega, J. L. *et al. N. Engl. J. Med.* **360,** 354–362 (2009).
5. Simon, T. *et al. N. Engl. J. Med.* **360,** 363–375 (2009).
6. Paynter, N. P. *et al. Ann. Intern. Med.* **150,** 65–72 (2009).
7. Lango, H. *et al. Diabetes* **57,** 3129–3135 (2008).
8. van Hoek, M. *et al. Diabetes* **57,** 3122–3128 (2008).
9. Lyssenko, V. *et al. N. Engl. J. Med.* **359,** 2220–2232 (2008).
10. Mihaescu, R. *et al. Genet. Med.* advance online publication doi:10.1097/GIM.0b013e3181b13a4f (2009).
11. Benjamin, E. J. *et al. Nature Genet.* **41,** 879–881 (2009).
12. Gudbjartsson, D. F. *et al. Nature Genet.* **41,** 876–878 (2009).
13. Gudbjartsson, D. F. *et al. Nature* **448,** 353–357 (2007).

NASA needs more money to track asteroid threats

NASA won't reach its mandated goal of tracking nearly all potentially hazardous near-Earth asteroids by 2020 with its current funding, according to an interim report published by the US National Research Council last week.

In 2005, the US Congress gave NASA a deadline of 2020 to detect, track and characterize 90% of near-Earth objects bigger than 140 metres. This is the size of objects thought to pose a significant risk if striking in urban areas, and at least twice the size of the object that levelled 2,000 square kilometres of Siberian forest in the 1908 Tunguska explosion.

The report, the final version of which is expected at the end of the year, says that the goal could be reached if more funding is found for future facilities, such as the Large Synoptic Survey Telescope or the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS). It also cites the importance of the large Arecibo radar telescope in Puerto Rico, which is well suited to characterizing asteroids once they are found.

World population will grow fastest in poorest areas

The world's least-developed regions will double in population between now and 2050, from 828 million to 1.66 billion, predicts the non-profit Population Reference Bureau, based in Washington DC, in data published last week. These 49 countries, 33 of which are in Africa, have the lowest incomes, highest economic vulnerability and poorest human-development indicators according to United Nations definitions.

The bureau's projections show that, over the same time period, the population of the world's more developed countries will creep up 7%, from 1.23 billion to 1.32 billion — fuelled mainly by immigration from less-developed regions. In the United States, however, more than half of the

expected growth will be due to births there; it has one of the highest fertility rates in the developed world.

Worldwide, the population could hit 9.4 billion in 2050, up from 6.8 billion today and topping 7 billion in the latter half of 2011. By 2050, India will overtake China as the world's most populous country.

Australia seeks carbon-reduction compromise

Following last week's defeat in the Senate of his proposed scheme for reducing carbon emissions, Australian Prime Minister Kevin Rudd was this week hoping at least to spur billions of dollars of renewable-energy investment.

As *Nature* went to press, a law requiring 20% of Australia's energy to come from renewable sources by 2020 was expected to make its way through parliament.

The energy bill was delinked from a much broader legislative package that would use a cap-and-trade scheme to limit greenhouse-gas emissions from industry (see *Nature* 458, 554–555; 2009). But that was voted down by opposition politicians on 12 August. Rudd intends to re-submit the contentious emissions-trading scheme to Senate for voting before the end of the year.

Innovation urged for water management in Asia

Asian irrigation systems require urgent reform, according to a report published on 18 August by the International Water Management Institute in Battaramulla, Sri Lanka, together with the Food and Agriculture Organization of the United Nations and other partner groups.

Asia contains 70% of the world's irrigated land, much of it watered by state-funded canal systems that were installed in the 1970s. But the systems are poorly maintained. If Asia is to meet the needs of its growing population, the report says, it needs more efficient, better-regulated

irrigation systems; more involvement by the private sector in managing them; and more education about, and investment in, watering programmes.

The report also suggests that governments should support and regulate extraction from groundwater aquifers by individual farmers, rather than condemn the practice.

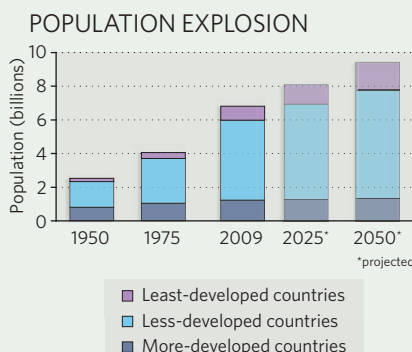For a longer version of this story, see http://tinyurl.com/news-2009-826

Canals in many Asian regions are in dire need of maintenance.

Conflict of interest and resignation at drug agency

The upper echelons of the US Food and Drug Administration (FDA) received unwanted attention last week. Daniel Schultz, head of the FDA's medical device division, resigned after allegations from employees that products were approved despite the concerns of agency scientists.

And it emerged that California-based Amphastar Pharmaceuticals had in April asked the FDA that Janet Woodcock, director of the FDA's Center for Drug Evaluation and Research, be removed from judging drug approvals of a generic blood-thinner because of her interactions with a competing firm, Momenta Pharmaceuticals, based in Cambridge, Massachusetts. Woodcock asked Momenta's co-founder, Ram Sasisekharan, to lead an FDA task force investigating tainted Chinese-made heparin last year, and she and Momenta scientists then co-authored medical journal articles on the topic. The FDA is looking into the complaints.

POPULATION EXPLOSION

Population (billions)

- Least-developed countries
- Less-developed countries
- More-developed countries

projected (1950, 1975, 2009, 2025, 2050*)

INT. WATER MANAGEMENT INST.

CLIMATE GLOOM

Reactions after slow progress at United Nations climate talks in Bonn, Germany, on 10–14 August

"If we don't have more movement and more consensus than we saw here, we won't have an agreement."
Jonathan Pershing, US lead climate negotiator.

"The best likely outcome in Copenhagen may be an interim agreement nailing down the basic architecture."
Elliot Diringer, vice-president of the Pew Center on Global Climate Change in Arlington, Virginia.

"My fear is that we sign another agreement that doesn't have any teeth."
Kevin Conrad, delegate from Papua New Guinea.

"It would be incomprehensible if this opportunity were lost."
Yvo de Boer, executive secretary of the United Nations Framework Convention on Climate Change.

Sources: *New York Times*, AP, Bloomberg, AFP

RACK AND FIELD

Ecologists have struggled to reconcile what they see in the lab and in the wild. But both views are needed to understand the effects of extinction, finds **Virginia Gewin**.

More than 100 streams run through Bradley Cardinale's laboratory. Some trickle, others gush. Some are home to one species of freshwater algae, others to eight. And they were all created to answer a simple ecological question.

Cardinale wants to know how the number of species living in a stream affects the quality of water that flows through it. The problem calls for controlled, replicable experiments, hence the artificial streams in his lab at the University of California, Santa Barbara. But he must also show that the results he sees in the lab apply to the water found in natural streams. So he is running a parallel experiment in the Sierra Nevada mountain range to compare how communities with naturally low and high species diversity clean up the streams that flow there (pictured above).

Tackling this question has taken Cardinale five years, in part because he has gone to great pains to include an aspect of reality in his lab experiment — water flow, which influences diversity. In the next stage he will have to quadruple the number of streams to analyse the effects of adding one algae-eater and its predator into the system. "The complexity of this work is invigorating and, at times, frustrating," Cardinale says. "But in an era when up to half the world's species hang in the balance, we must confront the complexity head on."

Within that complexity, hope Cardinale and others, lie answers to the question of whether the current wave of extinctions will just result in fewer pretty flowers and birds to look at, or whether it will mean poorer soils, more carbon in the atmosphere and the loss of billions of dollars' worth of 'ecosystem services' — natural processes that benefit humans, such as water purification, pollination and pest control. How to measure the impact of species loss on ecosystems and their services has been the subject of lengthy, inconclusive and sometimes bitter debate among ecologists. It has taken a long time to work out the best way to ask the question, with lab and field experiments often pitted against one another, and researchers agonizing over whether they should sacrifice biological reality for experimental purity.

> "The complexity of this work is invigorating and, at times, frustrating."
> — Bradley Cardinale

In recent years, ecologists such as Cardinale have realized that 'field with lab' is a more productive path than 'field versus lab'. Findings from such combined studies suggest that researchers have underestimated how important biodiversity is to ecosystem functioning, which is both vindicating and sobering to those who stress nature's practical value. The next step, they say, is to try to use this hard-won knowledge to guide efforts to save what's left, and to help restore natural places to make them useful to human needs. "We need to amass the science necessary to make predictions now, before more species go extinct," says Cardinale, "so that we can understand how many species must be restored to degraded ecosystems in order to regain functions like pest or disease control."

Variety: the spice of life

By the mid-1990s, most of the world seemed to believe that conserving the species in an ecosystem was important to maintaining its natural functions. In December 1993, countries that are party to the United Nations Convention on Biological Diversity committed to developing strategies to conserve and sustainably use biological diversity because, the convention argued, conservation would provide environmental

B. CARDINALE

and economic returns. Yet up until then, there had been little direct evidence of diversity's importance for environmental and human welfare. The debate swung backwards and forwards for decades, between ecologists' intuition that diverse ecosystems must be healthier than impoverished ones, and theoretical studies that suggested that as the number of species in a system grew, more extinctions would be expected[1]. With the convention, governments, along with the researchers they look to for guidance, had an urgent need to clarify the relationship between diversity and function, so that they would know what was important to conserve.

In 1994, two papers brought new experimental approaches to the question. Using a long-term study of grassland plots in Minnesota[2], ecologists David Tilman of the University of Minnesota in St Paul and John Downing, who was then at the University of Montreal, Canada, found that plots with more species were less affected by droughts, showing smaller reductions in biomass, and recovering to their pre-drought state more quickly. In the same year, Shahid Naeem and his colleagues at Imperial College London used the 'Ecotron' — a system of walk-in cells containing square-metre plots, each stocked with differing numbers of plant, herbivore, decomposer and predator species — to show that more diverse microcosms produced more biomass and consumed more carbon dioxide. A doubling or tripling of the number of species resulted in a roughly equivalent rise in productivity[3].

Tilman and Downing saw their results as support for the idea that the more species you have, the more likely a system will stay productive in the face of a drought or other crisis, because some species will be tolerant to the stress. Naeem's team suggested that the key lay in division of labour: a more diverse community is more fully able to exploit the spectrum of resources, such as light, soil nutrients or water availability. Either way, the implications for conservation were enormous. "If the number of species was important, then we had a responsibility to conserve every species," says Cardinale.

But many ecologists challenged the papers' conclusions. Until then, most scientists had believed that an ecosystem's function depended on whether it contained key species, with the total number of species having at most a minor effect. One of the fiercest critics, Michael Huston of Texas State University in San Marcos, thinks that the 1994 studies failed to exclude this possibility. "These experiments were designed, intentionally or not, to always show that more species have a higher level of ecosystem function," he says.

Huston argues that any relationship between diversity and function in those studies was probably due to a 'sampling effect'[4]; that is, higher productivity in diverse ecosystems could be explained by the greater probability that they would contain one or two highly productive species doing all the ecological heavy lifting. "No one disagrees that biodiversity influences ecosystem function," Huston says, "but is it the number of species or the properties of specific species?" Conservation efforts, he argues, shouldn't be based on the understanding that more species make for better ecosystem functions.

Some charged the researchers with blatant bias, accusing them of interpreting their studies so as to advance a political agenda for conservation by showing that biodiversity was economically valuable and important to human well-being. "Those comments were said in private conversations, and I personally found them astounding," says Tilman. He knew from the start that the study's conclusions would cause controversy, so he sat on his data for years, testing alternative hypotheses. "I didn't quite believe the results myself at first," he says. But Naeem, a postdoc at the time, and now at Columbia University in New York, was surprised that what he calls his "innocent experiment" caused such a stir. He says that his adviser, John Lawton, a community ecologist at Imperial College London, was

> "These experiments were designed, intentionally or not, to always show that more species have a higher level of ecosystem function."
> — Michael Huston

concerned that Naeem's experimental approach would jeopardize his nascent career. In ecology, says Naeem now, any experimental approach is liable to draw criticism. "The only thing that is real in science is what can be seen through a pair of binoculars. As soon as we start to mess with it, it becomes suspect," he says.

Even so, many ecologists spent the next decade messing with it. They built model ecosystems in an effort to control the myriad variables at work and to isolate the effects of species number. But in the quest for rigour, the experiments and the way they were manipulated became so unlike natural ecosystems that the goal — to uncover what happens when species go extinct — was sometimes missed. For example, it became customary for controlled experiments to randomize the addition or removal of species and environmental conditions, but such processes are not random in natural systems. Critics charged that ignoring the factors controlling biodiversity while studying the ecological effects of it rendered such studies irrelevant.

Heated debate

Temperatures rose as researchers took sides, some advocating the sampling effect, others arguing that efficient partitioning of resources among species was the key to increased productivity. "The debate had to happen, and it helped the community make progress to reconcile seemingly contradictory hypotheses, but it got a bit nasty," says Michel Loreau, an ecologist at McGill University in Quebec, Canada.

In 2000, Loreau and Naeem organized a meeting in Paris to encourage everyone to look at the same data and test hypotheses together. "We established a behaviour of confronting differing opinions," says Loreau. A consensus paper resulted[5], which proposed a quantitative method to parse out the effects of relationships between species versus Huston's sampling effect. It also prompted additional work to reconcile findings from theory, observation and experiment.

Most ecologists are now convinced that more-diverse ecosystems are generally more productive. To get the big picture, Cardinale led two meta-analyses, one of more than 100 experiments[6], the other of 44 (ref. 7), in which researchers had experimentally manipulated species in both artificial and natural aquatic and terrestrial systems.

Biodiversity in a box: 'Ecotron' chambers are used to measure effects of biodiversity on ecosystem productivity.

T. EVANS

The meta-analyses showed that diversity actually has a larger effect on productivity than anyone had previously documented. Mixtures of two or more species were more productive than the average monoculture 79% of the time, and mixed plots were 1.7 times more productive as those with single species[7]. The longer an experiment ran, the more pronounced the difference became, probably because ecosystem processes change as species settle into stable, complementary relationships. "We were stunned by the magnitude of diversity effects," Cardinale says. "If anything, we have underestimated the impacts of species extinction on ecosystem productivity."

The studies also revealed that although Huston's sampling effect is real, it accounts for just one-third of the overall increase in productivity. The other two-thirds is due to interactions between species and a better division of labour. A suite of different species seems to be able to partition resources in such a way that their own properties or specialties are performed more efficiently.

Like the sampling effect, however, this insight throws the spotlight back on the attributes of species. Some think that understanding how these properties, known as functional traits, contribute to ecosystem function is more useful in predicting the effects of extinctions than looking at species number. "We were so focused on the diversity of life," says Naeem, "we didn't look at the diversity of function."

Bridging the gap

Many ecologists, including Tilman and David Wardle, an ecologist at the Swedish University of Agricultural Sciences in Umeå, have been pursuing this approach all along. In 2005, for example, Wardle removed specific shrub species as well as specific functional traits, such as large roots, from 30 small islands in Swedish lakes. Some islands were affected very little and some a lot, showing how site-specific a trait's effects can be[8]. "Rather than simplify the system, we started with real ecosystems and measured what happens when different species are removed," says Wardle.

In recent years, ecologists have found other ways to bridge the gap between experimental clarity and biological reality. Experimenters have covered larger scales in space and time, and have expanded analyses across multiple levels of the food web, to include plants and animals, predators and prey. "It is possible to be rigorous and relevant at the same time,"
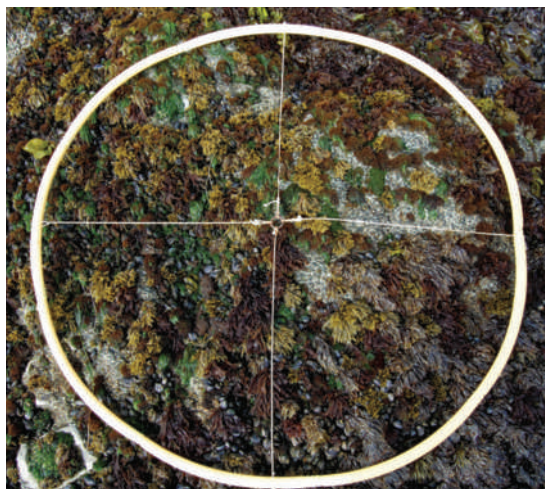
Rock and a hard place: seaweed studies compare the impacts of biodiversity in real tidal flats and artificial ones.

says John Stachowicz, a marine ecologist at University of California, Davis.

Like Cardinale, Stachowicz and his colleague Matthew Bracken, a marine biologist at Northeastern University's Marine Science Center in Nahant, Massachusetts, run parallel lab and field experiments. In 2008, they compared the effect of seaweed species richness on biomass in field plots on Pacific shores and in experimental tidal beds[9]. They found that more species meant a higher rate of biomass accumulation in the field, but not in the experimental chamber. They put this difference down to the fact that the chambers lacked the variability needed to pick up the effects of the division of labour, or the arrival or departure of species. They also found that the effects of biodiversity become stronger over time in the field, and suggested that many experiments did not run for long enough to see such effects.

> "It is possible to be rigorous and relevant at the same time."
> — John Stachowicz

Bracken is opening up a third front of investigation — *in silico*. His team has combined experiments with computer simulations of 3,500 combinations of seaweed species to model a larger number of scenarios than can be covered experimentally. The researchers found that with a realistic mix of species — which in nature, is controlled in part by wave exposure — more diversity meant more nitrogen uptake, whereas randomized species mixtures showed no such effect[10]. Most recently, he modelled a more realistic food web by adding snails to the mix, and found that their preference for sea lettuce (*Ulva lactuca*), the region's dominant seaweed species, increased overall plant productivity by allowing other species to grow[11].

Every small narrowing of the gap between experiment and reality reveals new things about ecosystem dynamics. But the gap may not be shrinking fast enough to reach firm conclusions — which may present a problem for conservationists who want answers now. "We're ten years away from telling policy managers which fraction of species to conserve," Cardinale says.

Cardinale is still optimistic — in a pessimistic sort of way — about ecologists' ability to provide practical advice. When he's not tending his 100 streams in the lab, he spends much of his time working in the field, wading in stream ecosystems that are losing species owing to residential development or nutrient pollution. "For many species and ecosystems, conservation efforts may have come too late," says Cardinale. But, he says, it may be possible to restore them, or to find another way to save or replace the ecological services they provide. "If we can nail down this field, we might be able to turn back the hands of time." ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

1. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1973).
2. Tilman, D. & Downing, J. A. *Nature* **367,** 363–365 (1994).
3. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. *Nature* **368,** 734–737 (1994).
4. Huston, M. A. *Oecologia* **110,** 449–460 (1997).
5. Loreau, M. et al. *Science* **294,** 804–808 (2001).
6. Cardinale, B. J. et al. *Nature* **443,** 989–992 (2006).
7. Cardinale, B. J. et al. *Proc. Natl Acad. Sci. USA* **104,** 18123–18128 (2007).
8. Wardle, D. A. & Zackrisson, O. *Nature* **435,** 806–810 (2005).
9. Stachowicz, J. J., Best, R. J., Bracken, M. E. S. & Graham, M. H. *Proc. Natl Acad. Sci. USA* **105,** 18842–18847 (2008).
10. Bracken, M. E. S., Friberg, S. E., Gonzalez-Dorantes, C. A. & Williams, S. L. *Proc. Natl Acad. Sci. USA* **105,** 924–928 (2008).
11. Altieri, A. H., Trussell, G. C., Ewanchuk, P. J., Bernatchez, G. & Bracken, M. E. S. *PLoS ONE* **4,** e5291 (2009).

J. STACHOWICZ

R. BEST

THE CHECK-UP

Patients checking in to the German Mouse Clinic will undergo the most sophisticated medical testing in the world. But, finds **Alison Abbott**, the waiting list is becoming a problem.

Each and every Thursday it is the same. At 10 a.m., the conference room at the Helmholtz Centre Munich starts to fill with top consultants and clinical researchers from the university hospitals in Munich, Bonn and Heidelberg. On the table is a stack — thicker than the city's telephone directory — of clinical data on a couple of hundred patients. The data are comprehensive — from blood analyses, to lung function, to metabolism. Some weeks there are brain and body scans in there too. After two or three hours of presentations and cross-speciality discussions, the group comes up with diagnoses they consider most likely.

It's a scene that could be imagined in any world-class hospital. But the patients involved are not human. This is the German Mouse Clinic, and the patients are genetically modified mice. The batteries of tests have been designed to identify every possible consequence of the animals' altered genes. And the results, the clinical scientists hope, will help them to work out what part each gene plays in cellular processes and in human disease.

Perhaps the most famous of the clinic's patients is the Foxp2 mouse, engineered to bear a 'humanized' version of the *Foxp2* gene, implicated in the evolutionary development of human speech and language. The gene, modified in the mice so that it contains two amino-acid substitutions that are normally present only in humans, changed the animals' ultrasonic calls and reduced their propensity to explore, but had no effect on organs outside the brain (W. Enard *et al. Cell* **137,** 961–971; 2009). "The negative results are as exciting as positive results," says Wolfgang Enard, a computational biologist at the Max-Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and creator of the *Foxp2* mouse. He was asked by the clinic to send a second batch of mice because staff couldn't believe the brain-specificity that they were seeing. The discovery that the humanized gene affects specifically the animal's brain and vocalization supports the idea that the amino-acid changes were important in human speech evolution.

Everyone in the world is welcome to send their mutant mouse to the clinic, which opened in 2001. And many are starting to do so. "We are getting more and more requests, and a waiting list is starting to build up," says geneticist Martin Hrabé de Angelis, who conceived and heads the clinic.

Geneticists estimate that around half of the 20,000 or so mouse genes have been altered by various means to make mutant strains. Some of these have been created expressly as animal models of disease or to get at a particular biological mechanism; others as part of systematic gene knock-out programmes such as the Knockout Mouse Project centred at the US National Institutes of Health (NIH) in Bethesda, Maryland. Many of these animals go through check-ups in the labs of the researchers that made them. But some have no obvious defects — and, for almost all of them, the rigorous 'phenotypic' analysis that the mouse clinic provides could reveal more useful data about their condition.

Unknown associations

From the point of view of its patients, the mouse clinic offers two months of comfortable living interrupted by 320 individual tests in 14 clinical areas ranging from neurology, to vision, to immunology. They have roomy accommodation, good food and a warm, pathogen-free environment.

Owners of the mutant mice complete a collaboration agreement with the clinic, and funding from the German research ministry allows the service to be offered free. The collaborators then provide the clinic with 80 mice, and testing begins. Some of the tests involve just being watched. The mice are placed in an 'open field' — a flat board with a camera suspended above it — while staff observe their breathing, how much they explore a new territory and with what enthusiasm.

Other tests can be more challenging for the mice. To have their blood pressure measured, they are placed in a restraining box with their tails hanging out. An air cuff, similar to the arm air cuff in a doctor's surgery, is wrapped around the tail. The procedure is done several times a day to follow the circadian changes in blood pressure. The mouse clinic has an array of sophisticated apparatus — for scanning the body in different ways, for example, or for assessing the structure and function of the heart using ultrasound — all scaled down to mouse size.
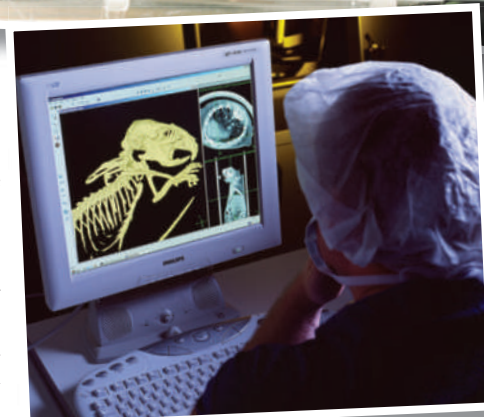
The work-up: mice are put through their paces on gripping (top), balance (middle) and coordination. Micro-computer tomography (bottom) is used to look for skeletal malformations.

At some point in their stay, the mice make an excursion to the neurology clinic. There they get to playfully hang on a rod while a scientist, by lightly pulling on their tails, gauges how strongly their paws can grip. The mice also balance on the 'Rotarod', which turns at increasing speed until they fall off, to gauge coordination. In the sealed perspex cube of the metabolic chamber researchers measure the amount of oxygen the mice breathe, the carbon dioxide they exhale and the heat they generate, to calculate basal metabolism. Everything the animals eat is weighed and every bit of faeces combed through, to determine how much energy they have absorbed from food.

Then, for those unlucky mice that are selected, comes pathology, histology and, as the web page euphemistically declares, 'chemical carcass analysis'.

All these tests and more are the fodder for the clinicians' Thursday discussions, to which the owners of the mice are invited. On the basis of the diagnoses, the clinical team may advise the owners to dig more deeply into particular phenotypes in secondary and tertiary screens that the clinic can perform. Something abnormal in antibody measurements, for example, might point to more complex allergy tests. Hesitancy in the open field, or a wobble on the Rotarod, might prompt advice to run brain tests, such as electroencephalograms, as well as sophisticated tests for memory, depression or anxiety, or sensory or motor function. Such screens are inevitably more time-consuming, more expensive and more complex than the primary screen.

Famous patients

Few other mouse facilities in the world perform such a breadth of phenotyping tests; hence few other mice on the planet are as closely scrutinized as the ones that enter this clinic. The level of data analysis and the generation of testable hypotheses also differentiates the German Mouse Clinic from screening offered elsewhere. Karen Svenson, who runs a mouse clinic at the Jackson Laboratories in Bar Harbor, Maine, says she hopes that "in the near future there will be an international effort to support worldwide clinic approaches", like that used by the Munich clinic.

More than 200 mutant mouse lines have been put through their paces at the German clinic since it opened. The Helmholtz Centre Munich supplies many of them from its programmes for creating mutant mice. One of these, the Aga2 mouse, is among those patients of the clinic with the most potential to influence medical practice. As a model for brittle bone disease, or osteogenesis

imperfecta, it carries a mutation in the *Col1a1* gene that is crucial in collagen production and mutated in the majority of people with this condition. It had been assumed that early death was usually a consequence of a stiffened ribcage. Unable to breathe deeply, patients succumb to multiple pneumonias.

Hrabé de Angelis was surprised when the clinical parameters thrown up in the primary mouse screen pointed in a different direction. "The blood pressure measurements weren't normal and there was something not quite right about the breathing," he says. The next round of screening revealed that the mutated *Aga2* gene actually causes direct, and fatal, damage to the lungs and heart. Joan Marini, an expert in osteogenesis imperfecta at the NIH, says that the insights offered have been "tremendous". "It will make clinicians more attentive to early preventative measures such as pulmonary exercise or bronchodilators, which could help keep the lungs elastic for longer," she says.

The clinic's waiting list is only likely to grow. Most biomedical researchers would like their mice to have more than one mutation per gene, to explore their different biological effects. On top of this, they are starting to recreate in mice the more complex types of genetic alterations that have been associated, however loosely, with human disease. Dusan Bartsch, a molecular

> "If we had had mouse hospitals earlier we could have advanced circadian biology by a decade."
> — John Hogenesch

biologist at the Central Institute of Mental Health in Mannheim, Germany, for example, engineers in mice specific genetic architectures that are frequently found in people with mental disorders such as schizophrenia. These include absent regions called 'microdeletions' and copy number variations, genetic regions that are repeated different numbers of times. "We are sending these mice to the clinic because there is just no way of predicting what other parameters could be affected by these complex gene changes," Bartsch says.

Hrabé de Angelis still wants to improve the range of testing. The phenotype expressed by any organism is dependent on the environment in which the organism finds itself. So the mouse clinic is piloting a new phase in which mice are tested in five different types of environment — what Hrabé de Angelis refers to as envirotypes. These include exposing the mice to stress, exercise or infection, giving them high-calorie 'cafeteria' diets and placing them in 'city' air polluted with diesel fumes.

But testing each mouse in the five envirotypes takes time, which will only make the waiting list even longer. To help keep pace with the number of mutant lines being created, Hrabé de Angelis is pinning hopes on winning long-term funding for the 'Infrafrontier' consortium, a European Union project to develop mouse functional genomics and, in particular, to upgrade other, smaller, mouse clinics around Europe so that they can perform at least the first round of screens. These are the Mouse Clinical Institute in Strasbourg, France, the UK Medical Research Council's Harwell unit and the Wellcome Trust Sanger Institute in Hinxton, UK, but others are now planned in Spain, Italy and the Czech Republic. There are already mouse clinics either open or planned in China and Japan.

Circadian biologist John Hogenesch from the University of Pennsylvania, Philadelphia, muses on how much faster his own field might have developed with the systematic phenotyping these clinics offer. The first mouse mutant with a disturbed circadian clock was made in 1997 (D. P. King *et al. Cell* **89,** 641–653; 1997) and since then others have been found in individual laboratories. "If we had had mouse hospitals earlier," says Hogenesch, "we could have advanced circadian biology by a decade." ∎

Alison Abbott is *Nature*'s senior European correspondent.

S. TÜPKE, MPI-EVA

CORRESPONDENCE

International peer review improved Irish research rankings

SIR — Your News story 'Italy outsources peer review to NIH' (*Nature* **459,** 900; 2009) highlights a problem common to many countries with a small population of research scientists. Ireland can be held up as a successful model in addressing this problem because, over the past eight years, funding agencies have moved to fully international peer review.

A few years ago, important research and development ventures were set up with a new infrastructure to attract talented people from abroad. The use of only Irish peer reviewers to allocate millions of research euros to a small number of universities could not stand up to the principles of objectivity, transparency and perceived fairness and would have led to conflicts of interest. Despite initial opposition, exclusively international review is now accepted; researchers want to be benchmarked internationally as well as nationally.

The typical process for research evaluation in Ireland is to consult four or five reviewers by mail for each proposal. Proposals are then assessed by a panel of invited experts, who meet in Ireland. Reviewers may be sourced through international funding agencies, or by letting applicants nominate experts themselves.

Some Italian scientists in your News story express reservations. They may well have a point, as US reviewers will probably not have any detailed knowledge of how research is conducted in Italy. One approach is to have nationals involved, either as observers or in a formal non-voting role. For example, the Irish Health Research Board (www.hrb.ie) organizes international mail reviews and panels, but the chair of each is Irish. They cannot participate in selection, but ensure that the correct procedures are followed and can explain the national research-funding policy. International panel members appreciate this local input, which helps them think outside their own national funding system.

Reviewing criteria often include the quality of the project, the researchers and their institutions, and the social and economic impact of the research. It is important that international reviewers focus on the quality of the first two, as the standing of institutions and the probable impact of a project can be harder for them to evaluate. Also, they should not get involved in detailed budgetary considerations, as these are strictly national.

Things have changed radically in Ireland's research over the past ten years. In 2008, the country appeared for the first time in a list of 'Top countries in all fields' (ranked by citations per paper; http://tinyurl.com/m5wdcl). We are now placed 19th, up from 36th place in 2003. I believe that international peer review played a significant part in this development.
Conor O'Carroll Irish Universities Association, 48 Merrion Square, Dublin 2, Ireland
e-mail: conor.ocarroll@iua.ie

Hispanic people start leaping over barriers to better jobs

SIR — Earlier this month, Sonia Sotomayor was confirmed as a member of the US Supreme Court, the first judge of Puerto Rican descent to be appointed. It is an encouraging result for both women and Hispanic people in the United States. Could it be that the barriers are at last crumbling for Hispanic people trying to reach high-ranking positions?

In the United States, people of Hispanic origin are notably underrepresented in science, technology, engineering and mathematics, and high achievers are rare. Although the proportion of Hispanic students enrolling in post-secondary courses in science, technology, engineering and mathematics in 2001 was the same as that of white or black ethnic groups, only 16% completed a bachelor's degree in these fields, compared with nearly 30% of white students (see http://tinyurl.com/ms9kpr). This picture has barely changed: according to the US Department of Labor, by 2012 about 15% of all US jobs will be filled by Hispanic workers, but only 1% will be at executive level.

With Hispanic people and their descendants expected to reach 30% of the total US population by 2050, closure of this gap comes none too soon.
Annelyn Torres-Reveron College of Pharmacy, Nova Southeastern University, 2250 Ave Las Americas, Ponce, 00717-9997, Puerto Rico
e-mail: at583@nova.edu

Whistleblowers at risk as science fails to correct itself

SIR — In his review of my book about Jan Hendrik Schön's fraud at Bell Laboratories, *Plastic Fantastic: How the Biggest Fraud in Physics Shook the Scientific World*, Martin Blume argues that the vigilance of whistleblowers is part of the natural corrective process of science ('Keeping up scientific standards' *Nature* **459,** 645–646; 2009). I disagree.

Blume's analysis fails to explain why whistleblowers so often find themselves working outside ordinary channels. In science today, the activities of searching for manipulated data, pursuing charges against colleagues and investigating others for misconduct are considered extra to normal scientific activity. Many suffer for speaking up, damaging their reputations. The fact that, in this rare case, the establishment did come around to the whistleblowers' position is not a reason to play down the risks they face, or the initiative and imagination needed to work around or take on reluctant scientific institutions.

Also, the correction of fraud often fails to happen through other recognized channels. Blume mentions the attempted replication of experiments by other laboratories and peer review, which were seriously wanting in the Schön case. Schön was only exposed when information about data irregularities previously noticed inside Bell Labs was passed along a chain of concerned scientists to independent external researchers. Whistleblowers only became involved after management failed to act effectively on internal concerns.

Science today can claim ownership of some corrective processes, including criticism of unrealistic research findings. Whistleblowing, sadly, is not among them.
Eugenie Samuel Reich Cambridge, Massachusetts 02139, USA
e-mail: eugenie.reich@gmail.com

See also Correction on page 957.

South Dakota school replies to sexual harassment claims

SIR — Your News story 'Sex scandal allegations surface at South Dakota school' (*Nature* **459,** 148; 2009) questions how the South Dakota School of Mines and Technology "will handle human-resource issues under its ... subcontract for the Deep Underground Science and Engineering Laboratory" in the light of alleged issues of sexual harassment. Your implication is unfair.

Although the school and its employees cannot comment upon the reasons for personnel actions, as director of human resources I wish to emphasize that we have no knowledge of any unresolved complaints, nor have we been apprised of any current concerns of the type alleged in the article.
Deborah L. Sloat South Dakota School of Mines and Technology, 501 E. St Joseph Street, Rapid City, South Dakota 57701, USA
e-mail: deborah.sloat@sdsmt.edu

OPINION

Biologists napping while work militarized

As researchers discover more agents that alter mental states, the Chemical Weapons Convention needs modification to help ensure that the life sciences are not used for hostile purposes, says **Malcolm Dando**.

In October 2002, Chechen rebel fighters held more than 750 people hostage at a Nord-Ost production in a theatre in Moscow. The siege was broken only after special military forces used what the Russian Health Minister, Yuri Shevchenko, later described as a mixture of substances derived from fentanyl — an opiate developed in the 1950s as an anaesthetic. Widespread relief that many of the hostages were saved was tempered by 124 of them being killed by the gas.

Chemicals with effects like those of fentanyl are often known as 'incapacitating agents'. These substances affect biochemical processes and physiological systems to produce a disabling condition such as unconsciousness, and in higher concentrations can cause death. With effects that last from hours to days, they are distinct from standard riot-control agents such as CS gas, which cause sensory irritation that disappears shortly after termination of exposure.

That Russian special military forces resorted to using fentanyl in Moscow is a possible harbinger of the wider militarization of advances in the biological sciences.

Designer weapons

Attempts to exploit benignly intended research for hostile purposes are not new. After the Second World War, the international medical community began to discover compounds that alleviated symptoms of mental illnesses such as depression and mood swings. These findings weren't accompanied by a good understanding of how the drugs worked. Nevertheless, they prompted nations to ramp up their efforts to find chemicals suited to military use. In fact, in 1959, the chairman of the UK government's secret Chemistry Committee of the Advisory Council on Scientific Research and Technical Development told his colleagues that the committee was "looking for agents which would produce, not cure, psychoses"[1].

Between the early 1950s and 1970s, researchers working in laboratories that eventually became the US Army Medical Research Institute of Chemical Defense studied chemical agents that affect the central nervous system. Indeed in 1961, the US military weaponized BZ — a drug that had originally been studied as a possible therapy for gastrointestinal diseases. BZ is one of a group of chemicals that act on the brain and can cause delirium; people exposed to it may fall into a stupor, struggle to speak, show poor coordination and have difficulty processing thoughts.

Despite the long-standing interest the defence industry has shown in drugs that alter people's physiological and mental states, a lack of knowledge has hampered attempts to use them. For example, by 1966, the US military had stockpiled munitions capable of delivering BZ, but its mode and site of action were poorly understood, and its effects varied widely from person to person. This and other problems led to its abandonment. The United States destroyed its stocks by 1990, several years before the Chemical Weapons Convention (CWC) entered into force in 1997.

Current biochemical threats range from lethal chemical agents to traditional and genetically modified biological agents. In general, biological agents such as anthrax cause governments the most concern. Only a few pathogens are suitable for military use, however. For example, smallpox could prove useful as a weapon because it is highly contagious; anthrax because it has a life cycle that involves the production of long-lived spores. The limited range of possibilities means that there is a good chance of developing countermeasures such as vaccines or antibiotics against these agents. Even if efforts are made to modify them — for example by introducing genes that encode antibiotic resistance — the problem of designing countermeasures is potentially surmountable because the range of effective manipulations that can be made is also limited.

But recent scientific and technological advances could transform the biochemical-threat landscape. Indeed, in 2003, military analysts from the Counterproliferation and Technology Office of the Defense Intelligence Agency in Washington DC predicted that emerging biotechnologies were likely to lead to a "paradigm shift" in the development of biological warfare agents[2]. They warned that it would soon become possible to engineer agents to target specific human biological systems at the molecular level.

This idea of identifying crucial biochemical pathways, and then designing compounds

> **"Some researchers are actively facilitating the development of new chemical weapons."**

SUMMARY
- Number of possible biological warfare agents on the rise
- Future targets could be emotions, immune responses or fertility
- Lack of engagement with problem among life scientists
- Weapons conventions need to catch up

to disrupt them is a leap from the traditional model of biological-agent development. It expands the options: there are likely to be thousands of potential molecular targets and numerous ways of disrupting each one.

Frontiers of concern

Concerns about this kind of expansion of biochemical threats have since been reiterated by scientific and medical communities. For example, in 2006, the US National Academies produced a report called *Globalization, Biosecurity, and the Future of the Life Sciences*. The authors argued that recent advances in our understanding of how bioregulatory compounds work, of signalling processes and of the regulation of human gene expression — combined with developments in chemistry, synthetic biology and in technologies such as nanotechnology — have "opened up new and exceedingly challenging frontiers of concern".

More recently, a 2008 US National Academies report entitled *Emerging Cognitive Neuroscience and Related Technologies*, similarly argued that in cases in which 'agonists' of a particular system have been found to enhance some cognitive trait, an 'antagonist' might be developed that could reduce it and vice versa. If dopamine agonists enhance attention, say, so dopamine antagonists might disrupt it. They also warned, among other things, that nanotechnologies could overcome the blood–brain barrier and "exploit existing transport mechanisms to transmit substances into the brain in analogy with the Trojan horse".

Some researchers are actively facilitating the development of new chemical weapons. For example, a research group from Pennsylvania State University in University Park has identified several drug classes as potential non-lethal agents or 'calmatives'[3], including benzodiazepines and α2-adrenoreceptor agonists, as well as individual drugs such
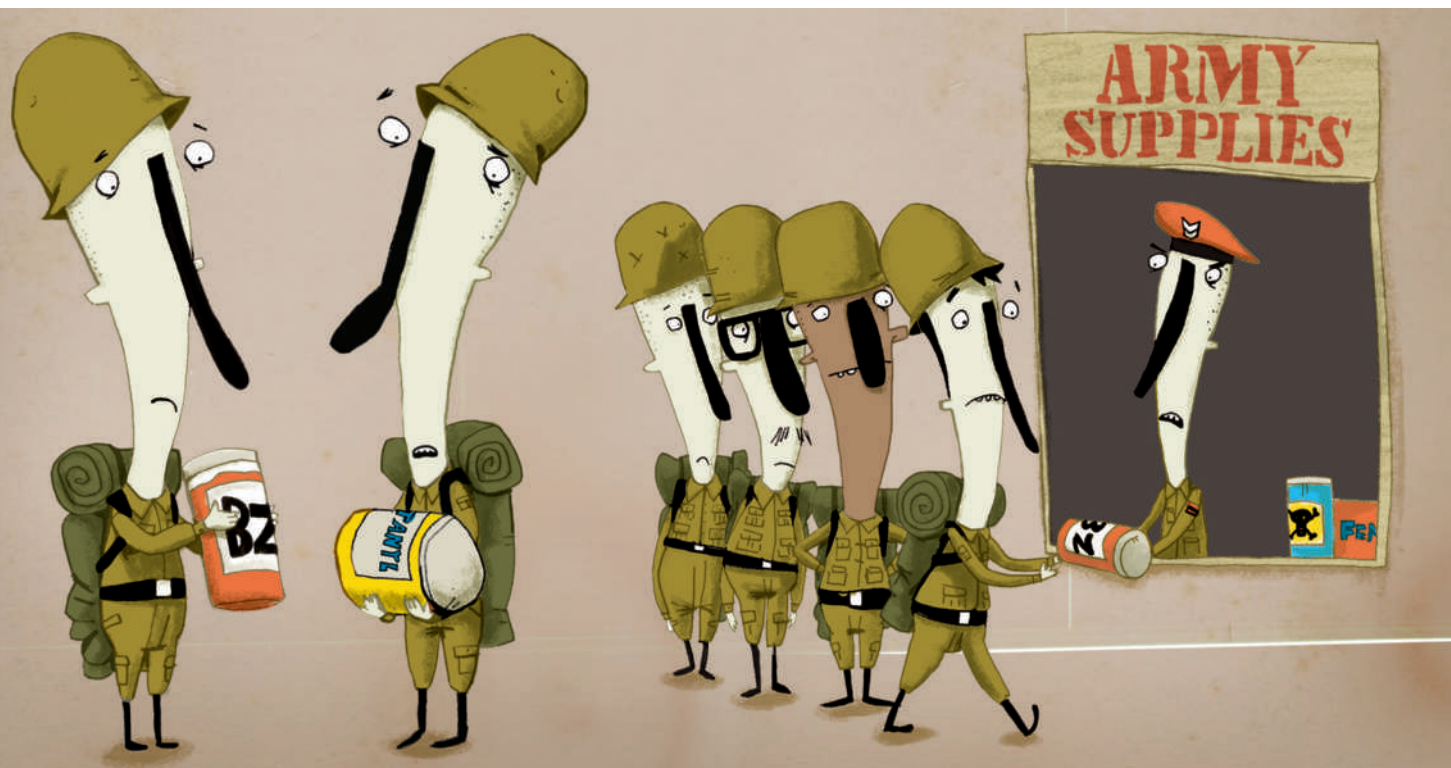
as diazepam and dexmedetomidine.

Similarly, at the 4th European Symposium on Non-Lethal Weapons in 2007, researchers from the Institute of Experimental Medicine and Charles University in Prague described the effects on macaque monkeys of combinations of drugs that produce a rapid loss of aggressive behaviour[4]. They argued that the drugs could be "used to pacify aggressive people during … terrorist attacks". The same researchers have also investigated methods of aerosol delivery to human volunteers.

Those who support the development of incapacitating agents often argue that using them in conflict situations stops people being killed. Historical evidence suggests otherwise. At the Nord-Ost siege, for instance, terrorists exposed to the fentanyl mixture were shot dead rather than arrested. Likewise, in Vietnam, the US military used vast quantities of CS gas — a 'non-lethal' riot-control agent — to increase the effectiveness of conventional weapons by flushing the Viet Cong out of their hiding places.

Blind to misuse

The lack of engagement with this issue among life scientists in general is alarming. Some companies are already marketing oxytocin on the back of studies showing that a nasal squirt of the hormone increases trust in humans. Even though the effectiveness of commercial sprays is doubtful, such research opens up the possibility of a drug that could be used to manipulate people's emotions in a military

context. Discussions with more than 2,000 practising life scientists in 13 countries over the past few years have taught me that few have considered such possible uses of their work.

There are problems with both the international conventions that protect us from the potential misuse of biological and chemicals research — the CWC and the Biological and Toxin Weapons Convention. The Biological and Toxin Weapons Convention, for instance, lacks an effective verification mechanism to ensure that nations are fulfilling their obligations.

Some have called for the CWC agreement to be amended to allow the use of novel incapacitating agents[5]. In the past 20 years, modern warfare has changed from predominantly large-scale clashes of armies to messy civil strife: think of Bosnia, Iraq and Afghanistan. The chemical agents described here are particularly suited to this style of warfare; it's not hard to find people in the military world who think they would be useful[5].

The CWC urgently needs modifying if it is to continue to help ensure that the modern life sciences are not used for hostile purposes. Most pressingly, the compatibility of the Convention with the development of non-lethal chemical agents for law enforcement needs careful scrutiny. Article II.9(d) states that "Law enforcement including domestic riot control" is exempt from the prohibition of the use of chemical agents. 'Law enforcement' could be taken by some to cover more than domestic riot control — which, in certain circumstances,

would make it legal for the military to use agents such as fentanyl.

In my opinion, all use of novel non-lethal agents such as fentanyl for law enforcement should be prohibited, or at least heavily restricted. If, instead, we sit on our hands we must accept that new incapacitating agents are just the beginning. We will be, as the British Medical Association concluded in its 2007 study, *The Use of Drugs as Weapons*, "knowingly moving towards the top of a 'slippery slope' at the bottom of which is the spectre of 'militarization' of biology" including "intentional manipulation of peoples' emotions, memories, immune responses or even fertility". ∎

Malcolm Dando is principal investigator for the Wellcome Trust project on Building a Sustainable Capacity in Dual-Use Bioethics, Department of Peace Studies, University of Bradford, Richmond Road, Bradford, BD7 IDP, UK.
e-mail: mrdando@bradford.ac.uk

1. Chemical Committee, Minutes of the 32nd Meeting, 5 March, 1959 (PRO/WO195/14637, 1959).
2. Petro, J. B., Plasse, T. R. & McNulty, J. A. *Biosecur. Bioterror.* **1,** 161–168 (2003).
3. Lakoski, J. M., Bosseau Murray, W. & Kenny, J. M. *The Advantages and Limitations of Calmatives for Use as a Non-Lethal Technique* (Penn. State Univ., 2000).
4. Hess, L., Schreiberová, J., Málek, J., Votava, M. & Fusek, J. Drug-Induced Loss of Aggressiveness in the Macaque Rhesus. *Proc. of the 4th European Symposium on Non-Lethal Weapons on Non-Lethal Weapons* Ettlingen, Germany, 21–23 May (2007).
5. Whitbred, G .N. T. *Offensive Use of Chemical Technologies by US Special Operations Forces in the Global War on Terrorism: The Nonlethal Option.* Maxwell Paper No. 37 (Air Univ. Press, 2006).

See Editorial, page 933.

Misadventures in the Burgess Shale

One hundred years after Charles Doolittle Walcott found a wealth of Cambrian fossils in the Rocky Mountains of British Columbia, **Desmond Collins** reflects on the bumpy road of their classification.

This August marks the centenary of Charles Doolittle Walcott's discovery of the Burgess Shale — the fossil bed in the Canadian Rockies that first opened mankind's eyes to the wealth of animals living in Cambrian seas 505 million years ago. The fossil bed itself is famous; for decades 'Burgess Shale life' was synonymous with 'Cambrian life'. Less well known is the full story leading to Walcott's discovery in 1909.

The finding of Burgess Shale fossils in Yoho National Park, British Columbia, was inevitable once the Canadian Pacific Railway ran its track through Kicking Horse Valley, between Mount Stephen and Mount Field in the 1880s. The first fossils were found on Mount Stephen by a carpenter helping to build a railway hotel, attracting the attention of Richard McConnell from the Geological Survey of Canada (GSC). In September 1886, 23 years before Walcott's more famous find, McConnell hit on the Mount Stephen trilobite beds: the first Burgess Shale fossil site.

McConnell collected a host of trilobites, a now extinct marine arthropod that is the trademark creature of the Cambrian. He also found two unusual specimens. Joseph Whiteaves, the GSC chief palaeontologist, described these in 1892 as the headless body of some sort of shrimp, noting that the appendages of the segmented body were not jointed, making it a sort of non-arthropod arthropod. He called it *Anomalocaris*, after this anomaly. Whiteaves unwittingly set the pattern for much misidentification and misclassification in Burgess Shale fossil research. The true nature of these fossils would not be revealed for 90 years.

The Darwin connection

Walcott first visited Mount Stephen in 1907 — the year that he was appointed secretary of the Smithsonian Institution in Washington DC — to study the Cambrian stratigraphy of the area. Two years later the Darwin centennial seems to have provided the serendipitous stimulus for his discovery of the Burgess Shale. Walcott was given an honorary doctorate at the University of Cambridge, UK, in June, as part of the 1909 celebrations. Afterwards he visited the British Museum (Natural History) in London. The keeper of geology there, Henry Woodward, had observed in a 1902 paper that Mount Field is part of the same massif as Mount Stephen

and "will no doubt yield the same Cambrian fossils". Presumably Woodward repeated this observation during Walcott's visit, because, on Walcott's return to the Rockies nine weeks later, he made straight for Mount Field.

On Saturday 28 August 1909, Walcott went up to Burgess Pass, next to Mount Field, to take photographs, and, according to his diary, "found the Stephen formation trilobite bed". His family followed him up to the site on 30 August. The story often told is that Mrs

Charles Dolittle Walcott discovered the most important, if not the first, Burgess Shale fossil beds.

Walcott's horse stumbled on the important 'discovery slab'. Instead, it seems that the man leading the packhorse train crossed over a rock that had slid onto the trail, and Walcott, concerned that his wife's horse might trip, moved it away. The next day, he and Mrs Walcott broke up the offending slab and found "a remarkable group of Phyllopod crustaceans". Walcott sketched three in his diary.

Walcott did not immediately realize the significance of what he had discovered. He spent just five days collecting at the site. But

he returned to find the source of the discovery slab in 1910, excavating the source rock layer and making a large collection of fossils.

In 1911, Walcott began to describe his finds. This was an unprecedented challenge. How to classify fossilized animals for which there is no known comparative material? This did not seem to bother Walcott. As a long-time trilobite expert, he began with the arthropods. One of these he named *Sidneyia inexpectans*, after his son Sidney, who had found the first specimen. Walcott portrayed it in a 1911 *National Geographic* article along with giant claws that had also been found at the site, with the caption: The King of the Cambrian World. Walcott never found such claws attached to *Sidneyia*; it seems that he simply associated the large, obviously arthropodan claws with the largest arthropod present. Like Whiteaves before him, Walcott got it wrong.

Stab in the dark

Walcott's next two papers, also published in 1911, were on four sea cucumbers, one jellyfish and some annelid worms. These descriptions revealed the exquisite preservation of Burgess Shale fossils. These were all soft-bodied animals, with no hard parts. Preservation of such creatures is extremely rare in the fossil record, and was unknown for the Cambrian.

Walcott's identifications did not fare well. Only one of the sea cucumbers is still thought to be a sea cucumber, and the jellyfish is now known to be part of an extraordinary arthropod. Of the 12 annelid worms, only one is still recognized as such. Walcott was on more familiar ground with his publication of arthropods in 1912 but, even here, many of the 18 genera he described are now known to belong to classes different to those to which he assigned them.

By the end of Walcott's final excavation in 1917, he had collected some 65,000 fossils, and observed: "This practically exhausts a quarry which has given the finest and largest series of Middle Cambrian fossils yet discovered in any formation in any country." Despite his obvious pride in making such a fabulous collection, Walcott's main goal continued to be to study the Lower Palaeozoic stratigraphy of the Canadian Rockies. Of his 18 seasons in the area, he spent 15 studying this, and only 3 (spread over 5 years) on the Burgess Shale fossils.
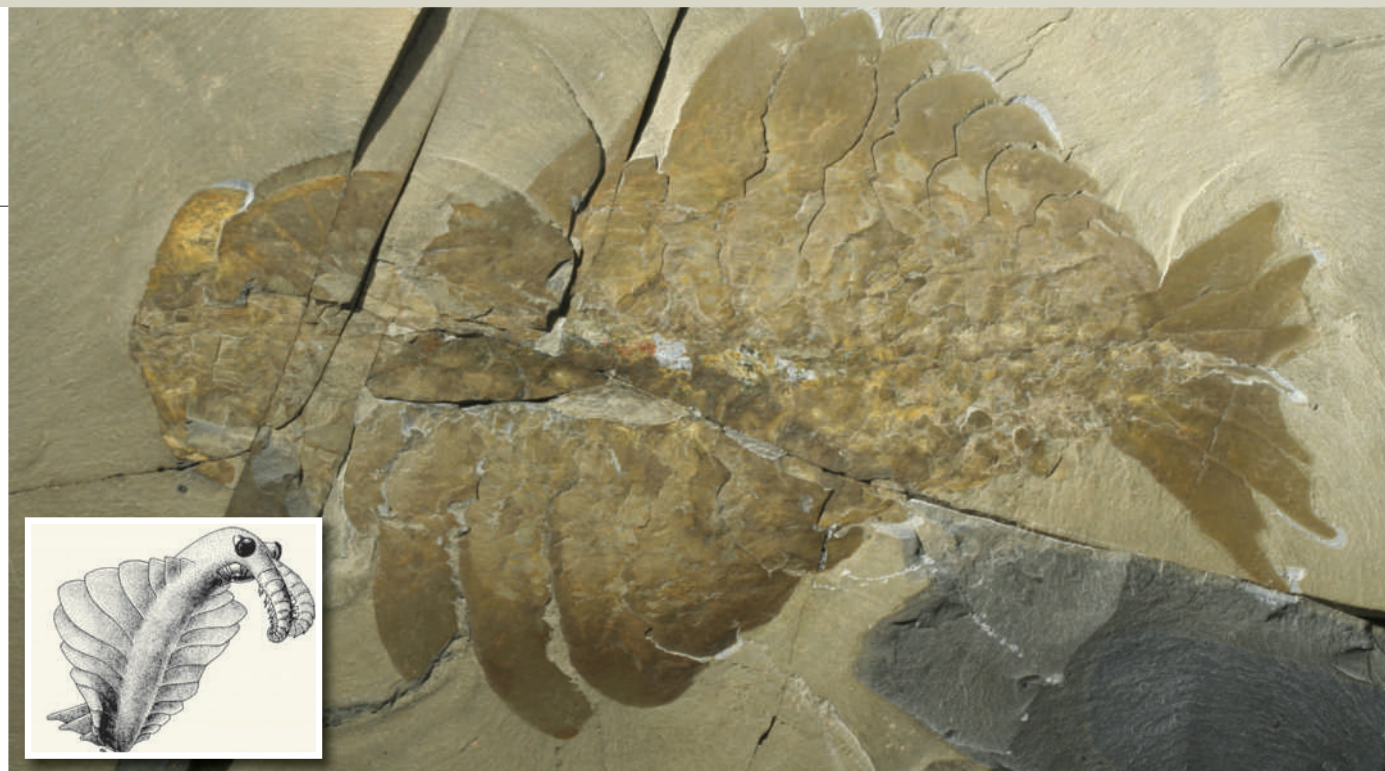
Anomalocaris, **once described as a shrimp, was revealed by Desmond Collins's Burgess Shale fossil find in 1991 (above) as a 'terrible crab' (reconstruction inset).**

After Walcott's death in 1927, his Burgess Shale fossils languished on the upper shelves of the Smithsonian collections. His third wife, Mary Vaux, was unwilling to have others rifling through them. This hiatus ended in the late 1960s when Harry Whittington of the University of Cambridge and his colleagues began a restudy of Burgess Shale animals, based mainly on Walcott's collections.

Whittington changed the world's perception of Cambrian life when he redescribed *Opabinia*, a bizarre looking animal with a frontal nozzle and five eyes. Walcott had classified it as an arthropod, but Whittington could find no appendages, jointed or not, so labelled it of "unknown affinities". This was the first time that anyone had questioned Walcott's implicit assumption that Cambrian animals belonged to groups alive today.

Weird wonders

As for the *Anomalocaris* 'shrimp' first described by Whiteaves, Whittington's student Derek Briggs recognized in 1979 that it was more likely to be an appendage of some larger creature. Soon after this, Whittington began to excavate from the enclosing rock what seemed to be part of one of these appendages in an unknown fossil. To his amazement, his preparation revealed two *Anomalocaris* appendages as frontal claws. Moreover, what Walcott had called a jellyfish in 1911 was revealed to be radiating jaws in a circular mouth at the front. This animal was like nothing ever seen before — "unlike other shrimps", as Whiteaves had written, was an understatement.

Whittington's work left *Anomalocaris*,

Opabinia and six others, mostly described by another of Whittington's students, Simon Conway Morris, unclassifiable in any known animal group. Stephen Jay Gould, in his 1989 book, *Wonderful Life: The Burgess Shale and the Nature of History*, took the next logical step: he suggested that the eight unclassifiable "weird wonders" belonged to animal groups that had become extinct.

Gould criticized Walcott for "shoehorning" his animals into known groups, so delaying the true understanding of the Burgess Shale animals. He attributed this to Walcott's conservatism and Presbyterian upbringing. To me, this is nonsense. None of Walcott's contemporaries, nor indeed the scientists who followed him, questioned Walcott's assumption that the Burgess Shale animals belonged to living animal groups; not until Whittington. True, many of Walcott's zoological assignments were wrong, but this led others to attempt to correct his mistakes. In effect, Walcott's misadventures provided the incentive for later revelations.

Since 1975, I have led 18 seasons of fieldwork and excavation in the Burgess Shale by Royal Ontario Museum parties. We greatly expanded the existing quarries, and discovered and excavated three new faunas, each with its own distinct composition. In 1991, we found the first complete *Anomalocaris*. For the first time, we could see what it actually looked like: a metre-long, fierce predator with large claws and radiating jaws. Two forms were described and assigned to a new class of arthropods, the Dinocarida — the 'terrible crabs'. The dinocarids preyed on animals in Cambrian seas

> **"Virtually all animal groups alive today were present in Cambrian seas."**

much as dinosaurs preyed on Mesozoic land animals. They change, radically, our view of the Cambrian way of life.

With a flood of new specimens coming from 12 seasons of excavation between 1983 and 2000, new forms are being described and old forms redescribed. Today, we have returned mostly to Walcott's practice of classifying Burgess Shale animals in living animal groups, but the groups are different. There are some extinct classes, such as the Dinocarida, but very few extinct phyla. Five of Gould's weird wonders have been classified, only one in a new phylum. This year, one of the conundrums from Walcott's time was solved — the claws of The King of the Cambrian World are now known to belong to *Hurdia,* a 'terrible crab'.

Additional Cambrian material is now coming from the Chengjiang fauna in China (particularly new chordates, the group that includes humans), and the Sirius Passet fauna in Greenland. Along with the Burgess Shale animals, they demonstrate that virtually all animal groups alive today were present in Cambrian seas.

Walcott's 1909 discovery was not the first, but the best Burgess Shale site. By his collections and publications, Walcott contributed more than anyone, before or since, to drawing back the curtain obscuring our view of the life of the Cambrian world. For this alone, he deserves the glory of this month's centennial celebrations. ■

Desmond Collins served as curator of invertebrate palaeontology and head of palaeobiology at the Royal Ontario Museum from 1968 to 2004. He is now retired.
e-mail: Suzanne.collins029@sympatico.ca
Further reading online at http://tinyurl.com/oxgo9k.

BOOKS & ARTS

Leading the fight against smallpox

Donald Henderson directed the World Health Organization's effort to eradicate the variola virus. His memoir is a lesson in managing complex projects and personalities, says **John Carmody**.

On 8 May 1980 the World Health Assembly declared that "the world and all its peoples have won freedom from smallpox". The assembly went on to express "its deep gratitude to all nations and individuals" who contributed to the success of its global vaccination programme, initiated in 1958.

One of those individuals was Donald Henderson, who directed the eradication campaign during its intensified period between 1967 and 1977. In *Smallpox: the Death of a Disease*, he recalls the personalities, politics and battles behind that resolution. Henderson skewers the declaration's self-congratulatory tone, quoting from Robert Burton's *The Anatomy of Melancholy* (1621): "When a thing has once been done, people think it easy; when the road is made, they forget how rough the way used to be." Moreover, after success they forget their former indifference in their zeal to claim a share of the credit.

The origins of the smallpox variola virus, around 12,000 years ago, are obscure. But its influence on world history has been profound — it has killed rulers in ancient Egypt, China, Japan, Africa and Europe. Its threat was challenged in 1798 by physician Edward Jenner's discovery that infection with the related cowpox virus conferred immunity to smallpox. But outside Europe, access to this vaccine was limited early on owing to its instability.

The geographical disparities in the incidence of smallpox were highlighted in 1947 in the *Weekly Epidemiological Record* published by the fledgling World Health Organization (WHO), which reported that smallpox incidence was increasing in Africa, Asia and the Americas, although declining in Europe. The World Health Assembly (WHA) in 1948 established a study group on the disease at its first formal meeting. At its eleventh meeting, the Soviet Union requested that the WHO "investigate the means of ensuring the worldwide eradication of smallpox" and "encourage" the production of adequate stocks of vaccine. But progress was desultory.

Field workers had a vital role in the global campaign to eliminate smallpox.

Then, in 1965, the 18th WHA resolved that global eradication of smallpox was a major objective of the WHO. This was mainly because of altered policy in the United States — its government did not want to seem weak in those cold-war years, and the expansion of air travel increased the risk of the disease's spread. Henderson was appointed as director of that global effort. Taking up his place in medical history, Henderson moved to Geneva, Switzerland, from his post as chief of the surveillance section of the Communicable Diseases Center (now the Centers for Disease Control and Prevention) in Atlanta, Georgia.

His easy narrative is convincing, in part because of his central role but especially because of his generosity towards the numerous other participants — he portrays the 'front-line troops' in the field as being even more important than those at WHO headquarters. His magnanimity makes his criticisms all the more trenchant and cogent.

Anyone can be blinkered to the real problems of the world when they observe them from a comfortable office in a prosperous society — and that is true of governments as well as individuals. Henderson's accounts of inefficiencies and hostility are dispiriting: the failure of one government to admit that a locally made vaccine was useless and thus dangerous; the insistence by another that home-manufactured vehicles must be exported to distribute that nation's aid, adding costs, delays and interruptions to the work; and the arrogance and obstruction of individual health officials.

The achievement — with dizzying statistics of the hundreds of millions of doses of vaccine that were distributed — becomes all the more admirable when Henderson describes the physical obstacles that were overcome. His relationships with people across many nationalities and backgrounds were a crucial factor. The imaginative recruitment of schoolchildren, for example, is particularly striking: they tend to know the goings-on in their communities and, unlike officials, can be relied on to say what is really happening.

There is a second act to Henderson's drama that is linked to the power politics that prevailed at the end of the cold war. In May 1980, the WHA urged "the discontinuation of smallpox vaccination" and resolved that no more than four WHO collaborating centres should hold and handle stocks of smallpox virus. This raised the question of which countries should hold

> **"Smallpox continues to hover as a dark and ominous cloud... It cannot be forgotten nor ignored."**

this potential bioweapon, one being the United States. Yet, in accord with the WHO resolution, the majority of the US population by this time had no immunity to smallpox. By then, Henderson had returned to the United States as a senior academic and later became adviser to President George H. W. Bush. Regrettably, here he is the epitome of discretion. He names few people, but does reveal his experience of the chilling intransigence and medical ignorance shown by the Pentagon in its negotiations with him over biological warfare policies.

Like all good stories, *Smallpox* recounts the deeds of heroes and villains, fools and sages. Henderson ends by declaring his pride in having been at the WHO at this crucial time. A better conclusion would have been the close of the penultimate chapter, in which his delight is tempered with a dash of reality: "smallpox continues to hover as a dark and ominous cloud as it has throughout the course of mankind's history. It cannot be forgotten nor ignored". ■

John Carmody is a medical scientist at the University of Sydney, NSW 2006, Australia, who writes and broadcasts scientific history.
e-mail: jcarmody@med.usyd.edu.au

Space for improvement at NASA

Mark Twain once said that "History doesn't repeat itself, but it does rhyme." Julianne Mahler, a political scientist at George Mason University in Fairfax, Virginia, makes a case for this observation in her examination of NASA's organizational responses to the losses of the space shuttles *Challenger* in 1986 and *Columbia* in 2003. She asks what NASA learned from the first accident, how that changed it and whether those changes were in effect when the second accident occurred. Given the high-risk nature of human space flight, was NASA a 'learning organization'?

Organizational Learning at NASA doesn't add to the extensive official record. But the book is a clear and insightful discussion of the factors that contributed to both accidents.

Mahler analyses information-processing structures, relations with contractors, political and budgetary pressures and organizational culture, including rivalries between NASA field centres. These four factors were cited as contributing causes to the shuttle failures through defects in mission management, safety monitoring and responses to schedule and budget pressures.

The book contrasts prescriptions from two organizational theories: 'normal accident' and 'high reliability'. The former states that some accidents are unavoidable in complex systems that are tightly coupled, such as nuclear power stations. Unexpected interactions among such a system's components can escalate a small problem into a major failure more quickly than human operators can respond. Thus a tension arises between the need for both centralized control, to ensure safe operations, and decentralized authority, to generate creative solutions to unexpected problems. These two goals conflict; yet society requires that complex systems such as air-traffic control operate safely. With great effort this can be achieved and centralization and decentralization can coexist; these are called high-reliability organizations.

High-reliability organizations seem ideal for managing a space mission. Their organization is hierarchical and centralized for routine decision-making. But when systems are stressed, other modes come into play. Cooperative relations emerge between front-line practitioners and senior officials; ranks fade as everyone focuses on fixing the problem and ensuring the mission's success. The organization responds to emergency situations that threaten major system failure with pre-planned procedures that are rehearsed and updated. Thus high-reliability organizations are highly self-conscious about learning, in order to respond to unexpected, non-routine events.

The Columbia Accident Investigation Board recommended that NASA adopt elements of high-reliability theory, although it observed that "neither High Reliability Theory nor Normal Accident Theory is entirely appropriate". But the histories of *Challenger* and *Columbia* show that NASA has not fully integrated learning into its organization. After *Challenger*, there was a greater willingness to communicate problems up the management chain and clear lines of authority and accountability were created. But they blurred over the years as a result of budgetary and organizational turmoil.

Only minor errors of interpretation or omission creep in to Mahler's analysis. For example, the congressional direction to use commercial software — when no suitable product existed — isn't mentioned in her discussion of NASA's difficulty in implementing a financial reporting system for project managers. This omission could lead readers to conclude that difficulties were solely due to NASA's shortcomings. But imposed requirements also contributed to decisions made.

Budget and schedule pressures came from Congress and the White House. And political pressures from congressional supporters influenced the creation of specialist 'lead centres', rivalries between which may have hindered communications. Thus it is not always obvious what lessons are being taught and by whom. If an organization is struggling to survive or to do too much with too little, the attention required for mission success may falter. As Mahler says, "NASA did learn in some cases, at some times, about some things." Unfortunately, in spaceflight, especially human spaceflight, that isn't good enough. ■

Scott Pace is professor of the practice of international affairs and director of the Space Policy Institute at George Washington University, Washington DC 20052, USA.
e-mail: space1@gwu.edu

NASA

The *Challenger* tragedy forced NASA to change its organizational culture.

No more fish in the sea

The End of the Line
by Charles Clover. Documentary directed
by Rupert Murray
See http://endoftheline.com

So Long, and Thanks for All the Fish: the quirky title of Douglas Adams's novel could turn out to be visionary. Our rampant exploitation of the oceans is such that by 2048 we may be facing a future with no fish. This is the message clearly conveyed in *The End of the Line*, a documentary starring environmental journalist Charles Clover, based on his book of the same name.

The film takes narrator Ted Danson, an actor who co-founded Oceana, one of the world's largest ocean-conservation organizations, from the North Sea to the coast of Senegal where humans are fishing species to the brink of extinction. If history has a lesson, it is that once we have gone too far, there may be no turning back. Despite a moratorium on cod fishing in Canadian waters since 1992, the stock has shown little sign of recovery there. Now, European cod stocks and numerous other fish species worldwide face the same fate, owing to a combination of technological efficiency and ineffectual policy.

Clover popularizes the work of fisheries scientists, such as Daniel Pauly, a marine biologist at the University of British Columbia in Vancouver, Canada, who have pieced together a picture of imminent catastrophe in the

Frozen stocks may soon be the sole supply of tuna.

global ocean. Among the first to recognize that fisheries' catch rates were in decline worldwide, Pauly discovered in 2001 that the phenomenon had previously gone unnoticed owing to systematic distortions in catch trends that were skewed by incorrect reports from countries with big fisheries. In 2003, Boris Worm and Ransom Myers of Dalhousie University in Halifax, Nova Scotia, reported that 90% of all large fish — including tuna, swordfish, marlin and cod — had been removed from the ocean since 1950.

The End of the Line is informative. Clover's reporting reveals that bluefin tuna — the endangered species with perhaps the most alarming plight in the ocean — is allegedly being bought and frozen in bulk by major corporations. Once ocean supplies run dry, the frozen fish could be sold at sky-high prices.

Clover's portrayal of the global fisheries

problem falls down on two counts: over-simplification and polarization. Although current fisheries policy is inadequate, much of it is based on science. Clover suggests, for example, that the practice of discarding — by which some 7 million tonnes of caught fish are thrown back into the sea each year — has arisen because fishermen simply do not want the species they have caught. But wasteful discarding is more often the consequence of a fisheries policy that is designed to prevent fishermen targeting juveniles and species outside of their allotted quota.

Clover is quick to point out the culprits of the fisheries crisis — slippery politicians, greedy fisherman, thoughtless consumers and big business — while making activists and scientists the stars of the show. But in adopting a tone of advocacy, with its inherent moralism, Clover isolates viewers and misses an opportunity to place this problem in context.

Overexploitation of fisheries is one part of the huge dilemma that humans face in an increasingly resource-limited world. We can seek sustainability, but we will not be able to diversify our consumption indefinitely. And climate change will decrease marine resources further. Those most affected will be the fisherfolk of developing countries, who make up 98% of people who are directly dependent on fisheries for their livelihood.

Despite its polemic hue, *The End of the Line* is an emotive portrayal of one of humanity's greatest challenges, and a stark reminder that early and effective action can make a difference. ■
Olive Heffernan is editor of *Nature Reports Climate Change*.

S. VIDLER/PHOTOLIBRARY.COM

An eye for evidence

The Scene of the Crime:
Rodolphe A. Reiss (1875–1929)
Elysée Museum, Lausanne, Switzerland
Until 25 October

To mark its centenary, the Institute of Scientific Police at the University of Lausanne, Switzerland, has released from its archives 120 crime-related photographs taken by its founder, a pioneer of forensic photography called Rodolphe Archibald Reiss. Now on show in Lausanne's Elysée Museum (Musée de l'Elysée), the exhibition comes with a warning: it is not suitable for sensitive people or children under 14.

In popular television series, the victims of murder tend to look reassuringly healthy.

But real death is unmistakable. It is both awe-inspiring and banal, and in this collection, often gruesome. There are images of the decapitated corpse of a man that was retrieved from the Lausanne–Geneva railway, along with his head; the exsanguinated body of a woman who botched a home abortion; and the old woman who was attacked so violently with an axe that her false teeth flew out, landing some distance away.

Few of the cadavers pictured are identified. The curators want you to look at them as Reiss did — objectively, with the aim of extracting the maximum information about the perpetrator and the circumstances of the crime. As for motive, you are left in the dark. Because the photograph is your only source of information, you enter into it fully, taking in every detail. In

doing so, you cannot help but notice the misery of the conditions in which almost all of these crimes were committed.

The French criminologist Alphonse Bertillon is usually credited with turning the haphazard practice of forensic photography into a systematic science in the late nineteenth century. But it was his younger colleague Reiss who developed and disseminated those methods, and who founded the world-renowned institute in Lausanne. Reiss helped to establish scientific police services in Russia and Brazil, and in 1914, he became the Serbian government's official investigator into the atrocities committed by the German and Austrian armies in the First World War. A contemporary wrote that he would have made an excellent model for Sherlock Holmes: tall and youthful well into his fifties, he had eyes that shone with intelligence, and a passion for deciphering signs and traces.

Unlike fictional detective work, however,

forensic photography does not lend itself to fame. As legal evidence, the photographs that Reiss took in the service of the law would have fallen foul of that same law had they been published at the time, and until now the images in this exhibition have remained inaccessible to the public. Now, the authorities have decided to release them, even though a grandchild of one of the victims — or of one of the perpetrators — could still wander innocently into the exhibition and have a fright.

Seeing the photos alongside Reiss's notes — on the importance of recording the precise position of the corpse, for example, or on people's susceptibility to autosuggestion when it comes to judging physical resemblance — you are struck by how modern his thinking was. At the time, when the science of photography was still developing, many of the Swiss houses to which he was called had no electricity. A section on the usefulness of tattoos in identification is a powerful reminder of this contrast: the naked torsoes of ex-soldiers are photographed decorated with

Rodolphe Reiss pioneered the use of forensic photography for recording crime scenes in the early 1900s.

Edwardian ladies sporting wide-brimmed hats, fully clothed and revealed only from the neck up.

The methods for reading a crime scene may have evolved since Reiss's time, but the reasons for committing murder and the information contained in those scenes have not. Were Reiss to come back to life today, he would no doubt quickly find his feet in any forensic laboratory. ∎

Laura Spinney is a writer based in Lausanne, Switzerland.

Earliest sketches of the Moon

Cosmos and Culture: How Astronomy Has Shaped Our World
Science Museum, London
Until 30 December 2010

Amid the commemoration of the Apollo landings, another lunar anniversary has just passed quietly. Four centuries ago, on 26 July 1609, the English scholar Thomas Harriot pointed his recently acquired 'Dutch trunke' at the Moon and drew what he saw.

He thus became the first person to record astronomical observations through a telescope, a few months before Galileo Galilei did the same. But unlike Galileo, who was struggling to feed a family on a lecturer's salary and angling for the patronage of the Medicis, Harriot had made his fortune and had no need of fame. He was already a distinguished mathematician — a renowned algebraist

Thomas Harriot's magnified map of the Moon.

who had corresponded with Johannes Kepler on the physics of rainbows and who had produced a body of work on the physics of motion.

By 1609 Harriot had other reasons to avoid the limelight. His patron, Henry Percy, Earl of Northumberland, was a cousin of one of the gunpowder plotters who had sought to blow up the English Parliament in 1605, and Harriot himself was imprisoned and interrogated for his connection. His papers were lost after his death in 1621 and not rediscovered until 1784, still at Percy's estate. So he remains little known.

Two of Harriot's drawings of the Moon, one of which shows its seas and craters, plus his notes on the moons of Jupiter, are on display in London's Science Museum until the end of 2010 as part of the exhibition *Cosmos and Culture*.

They are accompanied by a first edition of Galileo's *Sidereus Nuncius* of 1610, a 1496 edition of Ptolemy's *Almagest* and first editions of Nicolaus Copernicus's *De Revolutionibus Celestium Orbum* (1543), Kepler's *Astronomia Nova*

(1609) and Isaac Newton's *Principia* (1687). Other highlights include Chinese astrological figurines, made during the Tang dynasty in the first millennium BC, and the 2-metre-long telescope through which William Herschel discovered the planet Uranus in 1781.

The exhibition's subtitle is 'How Astronomy Has Shaped Our World'. But any compelling stories that could be told about the way that astronomy forges links between artisan and genius, amateur and professional, royal patron and government agency or pure science and popular culture are lost here through poor display. The fine Harriot drawings cannot be viewed close up, and labels are provided through a touch-screen display that is situated far from the pieces it describes.

Because everything is thrown together in an arbitrary fashion — those mighty books rub shoulders with a pile of sci-fi paperbacks, and the Tang-dynasty figures sit near an astronomical edition of the board game Monopoly — this exhibition is both a must-see and a missed opportunity. ∎

John Whitfield is a writer based in London.
e-mail: ja_whitfield@hotmail.com

Correction
In the book review 'Keeping up scientific standards' by Martin Blume (*Nature* **459,** 645–646; 2009), we inadvertently omitted the author's declaration of competing interests, submitted before he was engaged to write the review. This has been restored online at http://tinyurl.com/mwserz.

LORD EGREMONT/WEST SUSSEX RECORD OFFICE

Deepwater paddies in Bangladesh.

PLANT BIOLOGY

Genetics of high-rise rice

Laurentius A. C. J. Voesenek and Julia Bailey-Serres

When subject to flooding, deepwater rice survives by shooting up in height. Knowledge of the genetic context of this and other responses to inundation will be a boon in enhancing rice productivity.

Deepwater rice lives up to its name: this variety can outgrow slowly rising floodwaters of up to 4 metres in depth. On page 1026 of this issue, Hattori and colleagues[1] describe how they have identified two genes, *SNORKEL1* and *SNORKEL2*, that contribute to this spectacular elongation response.

Rice — the seed of *Oryza sativa* — feeds billions. Although productivity per hectare has more than doubled since the 1960s, a further doubling will be necessary to meet projected requirements by 2050 (refs 2, 3). More than 30% of Asian and 40% of African rice acreage is cultivated in either lowland paddies (15–50 centimetres deep) or deepwater paddies (depth of more than 50 cm). But lack of control of water depth in rain-fed paddies can be a serious problem: in some areas, water levels rise progressively during the growing season and can reach several metres; in others, flash flooding can fully submerge plants for days or weeks. High-yielding rice varieties cannot survive either extreme of inundation. As a result, some flood-prone areas are planted with traditional local varieties that display a remarkable capacity for flooding-induced elongation — of up to 25 cm per day — or that can tolerate submergence for up to 15 days. But the high-yielding varieties are typically five times more productive than these flood-tolerant plants.

Hattori *et al.*[1] pinpoint three chromosomal segments that regulate flooding-induced stem elongation in deepwater
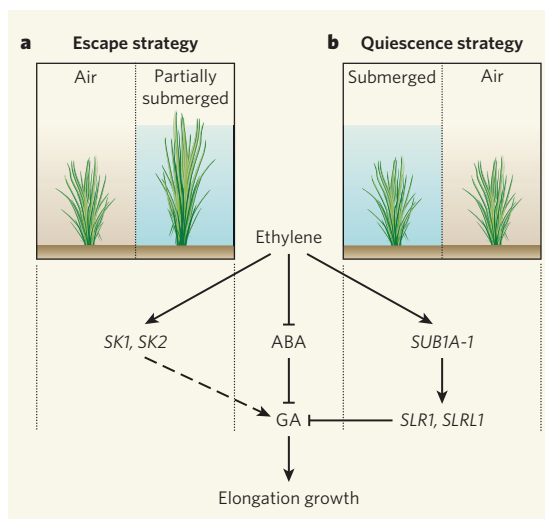
Figure 1 | Ethylene and flooding-tolerance strategies in rice. The elongation of stem and leaf cells is positively regulated by the hormone gibberellic acid (GA). In normal circumstances, a second hormone, abscisic acid (ABA), inhibits GA activity. When plants are submerged, ethylene, a gaseous plant hormone, accumulates owing to its slow outward diffusion in water. This promotes the breakdown of ABA, increasing the production of, or responsiveness to, GA, ultimately stimulating cell elongation. **a**, The escape strategy of deepwater varieties involves fast stem elongation to rise above the water level. Elongation growth and possibly GA accumulation or action are stimulated by transcription factors encoded by two ethylene-regulated genes, *SNORKEL1* and *SNORKEL2* (*SK1* and *SK2*) (dashed arrow)[1]. **b**, In the quiescence strategy of submergence-tolerant varieties, shoot elongation is suppressed so as to conserve carbohydrates and increase survival under flash-flood conditions. GA signalling and thus elongation are inhibited by the ethylene-induced action of a *SUBMERGENCE* gene (*SUB1A-1*) on the growth-inhibiting genes *SLENDER RICE-1* (*SLR1*) and *SLR LIKE-1* (*SLRL1*).

rice. When these regions were bred into a high-yielding, non-elongating rice (*O. sativa* subsp. *japonica*), the offspring gained the capacity to elongate in response to partial submergence. Hattori *et al.* focused on one segment of chromosome 12 that contributes 30% of the deepwater response, and identified a pair of genes, *SNORKEL1* and *SNORKEL2*, that stimulate underwater shoot elongation. Both genes belong to the large family of ethylene-response-factor (ERF) transcription factors, and are highly expressed in submerged stems. The *SNORKEL* genes are absent or non-functional in rice that lacks a deepwater elongation response, including *japonica* subspecies and *Oryza nivara*, one of two wild progenitors of domesticated rice. The *SNORKEL* locus seems to have been introduced from the other wild relative, *Oryza rufipogon* — there are populations of this species that have functional *SNORKEL* genes and display deepwater elongation.

Flooding triggers stem and leaf elongation by prompting a cellular increase in the volatile hormone ethylene, which promotes a decline in the hormone abscisic acid, followed by increased synthesis of, and response to, gibberellic acid[4,5] (Fig. 1). The interaction of gibberellic acid with its protein receptor induces the degradation of DELLA transcription factors, which normally limit cell elongation. Hattori *et al.*[1] found that the activation of *SNORKEL1* and *SNORKEL2* is induced by ethylene, which accumulates in submerged

organs, mainly as a consequence of the slow diffusion of gases in water as opposed to in air[4]. The action of the two *SNORKEL* genes apparently occurs downstream of ethylene but upstream of gibberellic acid (Fig. 1a). Further study is needed to see if *SNORKEL* directs the marked rise in bioactive gibberellic acid that Hattori *et al.*[1] see in submerged tissues of deepwater rice.

The *SNORKEL* locus is the second example of a multigene region in rice that encodes ERFs and regulates underwater growth. The *SUBMERGENCE-1* (*SUB1*) gene region, situated on chromosome 9, encodes two or three ERFs (SUB1A, SUB1B and SUB1C) that determine the response to complete submergence[6]. *SUB1* was identified in a traditional variety prized for its ability to endure more than two weeks of inundation. In contrast to the rapid stem elongation manifested in submerged deepwater rice, lines that have the *SUB1A-1* gene can survive the stress by limiting ethylene-activated elongation growth (Fig. 1b), thereby conserving precious carbohydrates for regrowth when the flood recedes[6,7]. This is accomplished by minimizing the decline in a DELLA protein (SLENDER RICE-1, SLR1) and a related non-DELLA protein (SLR LIKE-1, SLRL1) in submerged shoots[8]. Strikingly, therefore, ethylene can trigger antithetical outcomes: the promotion (SNORKEL) or suppression (SUB1) of underwater elongation.

Fast submergence-induced shoot elongation is also a characteristic of some wild plants that grow in flood-prone ecosystems. Together with other traits, such as the possession of aerenchyma (the snorkel-like conduits for fast gas diffusion)[9], underwater elongation determines the distribution and abundance of species in river floodplains and similar environments. Fast underwater growth in wild plants is also regulated by the interaction between ethylene and gibberellic acid, with downstream target proteins such as the cell-wall-loosening expansins also having a role[4]. Submergence-induced elongation is observed in plants inhabiting locations characterized by prolonged but shallow floods. By contrast, species found in places where deep transient floods occur limit shoot elongation during submergence[4] — like rice that adopts the quiescence strategy.

Wetland species that are quiescent during submergence also accumulate ethylene in submerged tissues and can respond to gibberellic acid under non-submerged conditions[4]. This suggests that ethylene-driven shoot elongation, and thus the ecological distribution of many wetland plants, is regulated by molecular components that control the production and response to gibberellic acid in an ethylene-dependent manner. We speculate that transcription factors, related to the SNORKEL and SUB1 ERFs, regulate submergence responses — and thus survival in flood-prone environments — across plant species. The presence, function, timing and tissue-specificity of ERF

expression could be targets of natural selection, and might thereby determine the remarkably varied growth responses mediated by ethylene at different doses and in distinct species[10].

Hattori and colleagues' delineation[1] of genes that control deepwater elongation has a practical edge: it provides a way to increase grain yields in areas prone to deep flooding by introducing the fast-elongation trait into high-yielding cultivars. This will complement the successful development of *SUB1* rice varieties that provide robust submergence tolerance in areas susceptible to flash floods[6]. Yields of rice cultivated in deep water may not reach those achieved in shallow paddies because carbon must be allocated to rapidly elongating underwater organs, and tall, grain-laden plants can topple if floodwaters recede. Nevertheless, the introduction of *SNORKEL* and *SUB1* into high-yielding varieties, using advanced breeding strategies, promises to improve the quality and quantity of rice produced in marginal farmlands. ■

Laurentius A. C. J. Voesenek is at the Institute of Environmental Biology, Utrecht University, Padualaan 8, 3584 CH, Utrecht, the Netherlands. Julia Bailey-Serres is at the Center for Plant Cell Biology, University of California, Riverside, California 92521, USA.
e-mails: l.a.c.j.voesenek@uu.nl; serres@ucr.edu

1. Hattori, Y. *et al. Nature* **460**, 1026–1030 (2009).
2. Khush, G. S. *Nature Rev. Genet.* **2**, 815–822 (2001).
3. Food and Agriculture Organization of the United Nations. *The State of Food and Agriculture 2008. Biofuels: Prospects, Risks and Opportunities* (FAO, 2008).
4. Bailey-Serres, J. & Voesenek, L. A. C. J. *Annu. Rev. Plant Biol.* **59**, 313–339 (2008).
5. Kende, H., van der Knaap, E. & Cho, H.-T. *Plant Physiol.* **118**, 1105–1110 (1998).
6. Xu, K. *et al. Nature* **442**, 705–708 (2006).
7. Fukao, T. *et al. Plant Cell* **18**, 2021–2034 (2006).
8. Fukao, T. & Bailey-Serres, J. *Proc. Natl Acad. Sci. USA* **105**, 16814–16819 (2008).
9. Colmer, T. D. *Plant Cell Environ.* **26**, 17–36 (2003).
10. Pierik, R. *et al. Trends Plant Sci.* **11**, 176–183 (2006).

CHEMICAL PHYSICS

Electronic movies

Marc Vrakking

Strong laser fields can tear an electron away from a molecule, leaving a hole in the electronic wavefunction that races through the molecule. The ultrafast motion of such a hole has been traced at last.

The first event in a light-induced reaction is that the electrons in the irradiated molecule respond to the incident light. This response occurs on attosecond timescales (1 attosecond is 10^{-18} seconds), and even the fastest time-resolved measurements have been unable to probe the dynamics of the process. But on page 972 of this issue, Smirnova *et al.*[1] provide a glimpse into the fascinating world of attosecond electron dynamics, in effect revealing the opening frames of a movie that records the rearrangement of electrons when a molecule is ionized by a laser.

The authors make use of the fact that molecules exposed to intense laser fields emit soft (low-energy) X-rays, in a process that is generally referred to as high-harmonic generation (Fig. 1a). In this process, an electron is extracted from a molecule by the laser, leaving behind a molecular ion. The electron is then accelerated in the laser's oscillatory electric field, before recombining with the molecular ion[2]. The recombination results in the emission of an X-ray photon. Large numbers of (typically near-infrared) photons are thus converted into soft X-ray photons, which have frequencies that are multiples (harmonics) of the source laser's frequency.

High-harmonic generation has recently attracted great interest for at least two reasons. First, the soft X-ray photons are not emitted continuously, but in short bursts that have

attosecond durations[3,4]. These attosecond pulses can be used to investigate fundamental questions about the interactions of light with matter. The number of research groups working in the field of harmonics-based attosecond science is thus growing exponentially, making it one of the most vibrant areas of laser physics and chemistry.

Second, harmonics contain a wealth of information about the atoms or molecules from which they are generated. Experiments have already suggested that high-harmonic generation allows information about molecular orbitals[5] and molecular structure[6] to be retrieved. The experimental and theoretical work reported by Smirnova *et al.*[1] represents a new milestone in this respect. The authors argue that high-harmonic radiation carries information about the amplitudes and phases of electron 'hole states' that are formed during the ionization event; holes are positively charged 'quasiparticles' formed by the absence of an electron. High-harmonic radiation can provide insight into hole migration in the ion during the timescale of up to several femtoseconds (1 femtosecond is 10^{-15} seconds) when the ionized electron is away from the ion, before recombination occurs.

Smirnova *et al.* used the fact that intense laser fields can force molecules to point in any desired direction — a process known as dynamic alignment[7]. By studying the interference
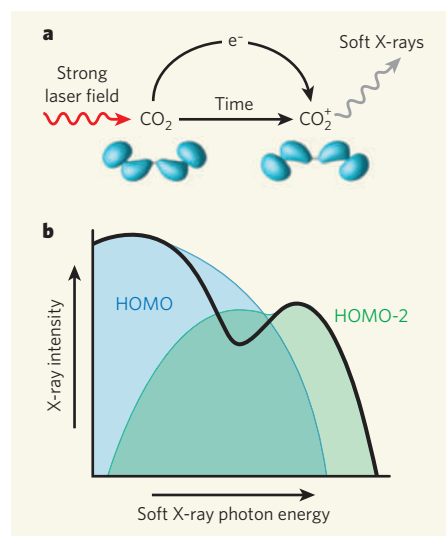
Figure 1 | High-harmonic generation. a, In high-harmonic generation, an electron is extracted from an atom or molecule, accelerated in a laser field, and directed back towards the resulting ion; soft X-rays are emitted during the course of electron–ion recombination. Smirnova *et al.*[1] show that the electronic wavefunction left behind by the ionization of CO_2 is different from the one encountered by the electron when it recombines with the CO_2^+ ion. Wavefunctions are shown in blue. **b**, The existence of different electronic states of the CO_2^+ ion gives rise to the formation of a minimum in the combined intensity of the X-ray emissions, caused by destructive interference of contributions from the different states (in this case, the HOMO and HOMO – 2 states). Blue and green areas are the X-ray emission spectra of the HOMO and HOMO – 2 states; the black line indicates the sum of these spectra after they have interfered destructively.

patterns of harmonic emissions from aligned and non-aligned carbon dioxide molecules, Smirnova *et al.* obtained the amplitude and phase of the harmonics as a function of the laser intensity, the alignment angle and the harmonic order (the multiple of the source laser's frequency that defines each harmonic frequency). In this way, the authors provided compelling evidence that, on ionization, the CO_2^+ ion is formed in both the ground state of the ion (corresponding to the removal of an electron from the highest occupied molecular orbital, known as the HOMO) and the second excited state (corresponding to removal of an electron from a lower-energy orbital known as HOMO – 2). This challenges the widely held notion that strong-field ionization processes remove electrons only from the HOMO. A similar conclusion was recently reached in a study of harmonic generation from nitrogen molecules[8].

If more than one ionic state is populated in the ionization, this suggests that the electronic wavefunction of the molecular ion evolves between the moment of ionization and the moment of recombination, when the hole vacated by the electron moves. During recombination, the structure of the hole is encoded in the harmonic spectrum, which thus in effect

provides a 'movie' of the hole's development. By working back from the information in the spectrum — playing the electronic movie in reverse — Smirnova *et al.* determine that the original ionization must have occurred by a 'tunnelling' process. Tunnelling is one of two models often used to describe the ionization of atoms or molecules influenced by a strong laser field.

This work[1] also sheds light on a fascinating controversy that has created quite a stir in the high-harmonics community. In 2002, the theoretical physicist Manfred Lein proposed[9] that harmonic generation from aligned molecules is influenced by the wave-like nature of the returning electron. He predicted that destructive interference could occur between harmonic emissions that originate from different atomic centres in a molecule, analogous to that observed in interfering light waves by Thomas Young in the classic double-slit experiment of 1801. The interference minima in harmonic spectra should depend only on the angle of alignment between the molecular axis and the electric field of the source laser beam, the wavelength of the returning electron and the internuclear distances of atoms in the molecules. Because the first two variables can be determined or controlled, this would allow the internuclear distances to be calculated. Remarkably, however, experiments in Tokyo[6] and Milan[8] observed interference minima for aligned CO_2 molecules at substantially different harmonic orders. Surely, the structure of CO_2 in these two cities could not be different?

Smirnova *et al.*[1] now reconcile these conflicting observations[6,8] by suggesting that the interference minima in the harmonic spectra of CO_2 have primarily a dynamical rather than a geometrical origin. They argue that the harmonic spectra consist of overlapping contributions from ionization of the HOMO and HOMO – 2 states. These two contributions interfere with

each other, creating a minimum at a harmonic order that is determined by the intensity of the laser used in the experiment (Fig. 1b). The authors present a convincing comparison of numerical simulations of harmonic spectra with experimental data over a wide range of laser intensities, and show that their explanation might account for the different results of the Tokyo and Milan experiments.

Does this mean that the structural interpretation of minima in harmonic spectra — and hence the prospect of using such spectra to determine molecular structures — goes out of the window? Not quite yet. Two recent papers[10,11] have argued that interference patterns can have a structural origin. This proposal is based on the experimental confirmation of a distinctive feature, known as a phase-jump, on either side of the interference minimum. Intensive research thus continues around the world, aiming to disentangle geometrical and dynamical effects in harmonic generation and searching for conditions under which the two can more easily be separated. The ultimate outcome may well be that high-harmonic generation leads to interference minima that have both a dynamical and a geometrical origin. For now, the jury is still out. ∎

Marc Vrakking is at the FOM Institute for Atomic and Molecular Physics (AMOLF), Science Park 113, Amsterdam 1098 XG, the Netherlands.
e-mail: m.vrakking@amolf.nl

1. Smirnova, O. *et al. Nature* **460,** 972–977 (2009).
2. Corkum, P. B. *Phys. Rev. Lett.* **71,** 1994–1997 (1993).
3. Paul, P. M. *et al. Science* **292,** 1689–1692 (2001).
4. Hentschel, M. *et al. Nature* **414,** 509–513 (2001).
5. Itatani, J. *et al. Nature* **432,** 867–871 (2004).
6. Kanai, T. *et al. Nature* **435,** 470–474 (2005).
7. Rosca-Pruna, F. & Vrakking, M. J. J. *Phys. Rev. Lett.* **87,** 153902 (2001).
8. Vozzi, C. *et al. Phys. Rev. Lett.* **95,** 153902 (2005).
9. Lein, M. *et al. Phys. Rev. Lett.* **88,** 183903 (2002).
10. Boutu, W. *et al. Nature Phys.* **4,** 545–549 (2008).
11. Zhou, X. *et al. Phys. Rev. Lett.* **100,** 073902 (2008).

NEUROSCIENCE

Activity acts locally

Jonathan B. Demb and Marla B. Feller

How does neuronal activity affect the development of neural circuits? Work on the retina shows that blocking activity at the synapses between neurons reduces local synapse assembly without affecting global cellular structure.

Nervous system function depends on the organization of underlying neural circuits — groups of neurons whose ability to perform specific functions depends on an organized pattern of intercellular communication at neuron junctions (synapses). The importance of synaptic activity in the assembly of circuits during development has been a long-standing debate in neurobiology: a dominant model in the field[1,2] suggests that the amount of neurotransmission at synapses during maturation

of the nervous system affects the large-scale arrangement of neural circuits, and even the structure of their component neurons. On page 1016 of this issue, Kerschensteiner *et al.*[3] present exciting work that revises this model. The authors show that, in the retina of mice, excitatory synaptic activity determines the density of synapses at individual neurons (a local effect), but does not affect cellular structure on a global scale.

Synaptic organization has been studied
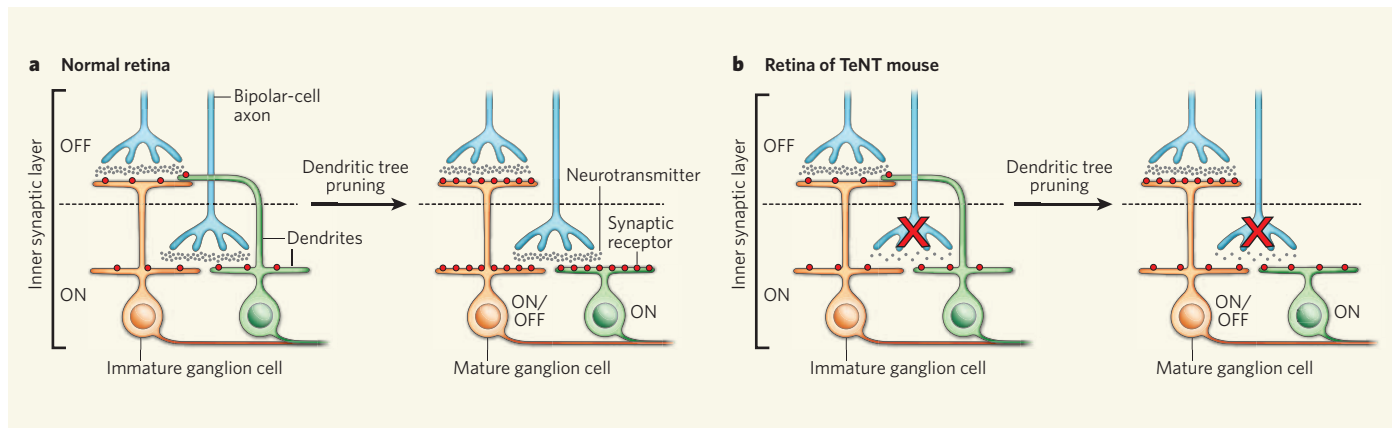
Figure 1 | Blocking synaptic activity in the retina has local effects. a, The inner synaptic layer of the retina is divided into two halves, in which ganglion-cell dendrites receive synaptic input from either OFF or ON bipolar-cell axon terminals. An immature ganglion cell that is destined to be bistratified (orange) maintains dendrites in both layers and matures into an ON/OFF cell. An immature ganglion cell that is destined to be monostratified (green) initially extends dendrites into both halves of the synaptic layer. During development, some dendrites are pruned and, at maturity, the monostratified ganglion cell makes contact with the axons of only one class of bipolar cell (ON, in this case). The density of synaptic receptors (red dots) increases during development. **b,** In the TeNT mouse studied by Kerschensteiner *et al.*[3], synaptic neurotransmitter release from ON bipolar cells is inhibited (red cross). ON/OFF and ON ganglion cells stratified normally in this mouse. For the bistratified ON/OFF cells, normal synaptic release in the OFF layer resulted in the normal increase in receptor density, whereas absence of synaptic release in the ON layer resulted in low receptor density, as in the ON cell. Thus, excitatory synaptic activity was required for the development of local synapse density but not for global cellular morphology.

extensively in the mammalian retina, a tissue comprised of about 15 discrete circuits, which can be distinguished both functionally and anatomically[4]. Retinal circuits are activated by photoreceptors (rods or cones), which transform light energy into neural signals and make contacts with both excitatory interneurons (bipolar cells) and inhibitory interneurons (horizontal cells) that process information within the retina. Subsets of bipolar cells and another class of inhibitory interneuron (amacrine cells) converge selectively onto one of about 15 types of ganglion cell. Long, thread-like projections, or axons, from these output neurons form the optic nerve and travel into the brain. The circuit of each ganglion-cell type determines its specific sensitivity to contrast, colour, motion and other features of the visual input.

A fundamental organizing principle of the retina is its division into ON and OFF pathways[4]. Bipolar cells are excited by either increases (ON cells) or decreases (OFF cells) in light; the axons of ON and OFF bipolar cells terminate in different sub-layers of the inner synaptic layer of the retina (Fig. 1a). Here, they release the neurotransmitter glutamate, which crosses synapses and activates excitatory receptors on short processes called dendrites that project from ganglion cell bodies.

Early in postnatal development in the mouse, before eye opening, ganglion cells extend dendrites to make contact with bipolar cells in both the ON and OFF retinal sub-layers. At maturity (three weeks of age), some ganglion cells remain bistratified, with dendrites in both layers, and thereby receive a mixed ON–OFF input[4]. Others lose their dendrites in either the ON or OFF sub-layer, maturing to become monostratified ganglion cells, which make contact with only one type of bipolar cell (ON or OFF) (Fig. 1a). Hence, the category of

a ganglion cell's response depends both on the morphology of its dendrites, which determines potential synaptic partners, and the presence of synapses, which enables actual communication with those partners.

The pruning of dendrites from ganglion cells as they become monostratified is thought to be induced by differences in excitatory synaptic activity, with the elimination of synapses (and eventually dendrites) from less-active inputs[2]. Kerschensteiner and colleagues[3] re-examine this hypothesis by generating mice with imbalances in synaptic activity at retinal ganglion cells. The ON bipolar cells in these mice were induced to express tetanus toxin (TeNT), which cleaves a protein needed to release neurotransmitters at synapses[5]. So, although the ON bipolar cells are structurally normal, synaptic neurotransmitter release from these cells is markedly reduced, and ON ganglion cells lack almost all excitatory synaptic input, both spontaneous and light-evoked.

Kerschensteiner *et al.* found that blocking synaptic activity of ON bipolar cells altered local ganglion-cell structure. The authors used confocal microscopy of fluorescently tagged proteins associated with glutamate receptors to visualize synapses in ganglion cells. They found that these receptors were in the correct location — near the silent ON bipolar-cell terminals — but that their number was halved. Furthermore, they found that the reduced density of synapses did not result from increased synapse elimination, but rather from a reduced rate of synapse formation between ON bipolar cells and ON ganglion cells.

Surprisingly, the reduced number of synapses in the TeNT mouse did not alter the stratification of ganglion cells. Indeed, in mature mice, dendrites of ON ganglion cells made contact with axon terminals of the silent

ON bipolar cells — somehow, the dendrites of ON ganglion cells found the correct target layer despite the lack of ON bipolar-cell synaptic activity. Moreover, bistratified ganglion cells also had normal dendrite morphology despite the imbalance in synaptic activity (active neurotransmitter release from OFF bipolar-cell synapses, but not from ON synapses) (Fig. 1b). Thus, ON and OFF bipolar cells that converge on the same ganglion cell apparently do not coordinate their synapse number so that synapse density is balanced. They also don't seem to compete for 'postsynaptic real estate' — if they did, the density of synapses at dendrites in contact with the active OFF bipolar cells would have increased in TeNT mice compared with normal mice.

Kerschensteiner and colleagues' results[3] raise an interesting question: why have previous studies shown that manipulations of synapse activity alter dendrite stratification whereas the TeNT manipulation did not? In cats and ferrets, intraocular injections of a drug known to hyperpolarize, and thereby functionally inactivate, ON bipolar cells disrupted stratification of ganglion-cell dendrites[2,6]. In mice, the normal stratification of ganglion-cell dendrites was delayed by altering spontaneous activity during development and was disrupted by depriving the retina of normal light input[2].

One possible explanation is that the TeNT mouse has retained several forms of synaptic activity. For example, OFF bipolar-cell synapses were apparently functioning normally, and some of their neurotransmitter could have reached ON ganglion cells through synaptic 'spill-over'[7]. Furthermore, the inhibitory interneurons of the TeNT mouse were presumably functional, and their synaptic release might have influenced dendritic refinement during development.

A second possibility is that manipulations in previous studies[2,6] had effects beyond reducing signalling through the ON bipolar-cell pathway. For instance, these previous manipulations may have unexpectedly affected the expression of molecules that guide the development of global cellular structure. Indeed, the layered arrangement of axons and dendritic processes in zebrafish[8] and chick[9] retina depend mostly on expression of adhesion molecules and relatively little on neural activity. However, it is difficult to exclude a role for synaptic activity in mammalian retinal development because the effects of previous experimental manipulations[2,6] on activity have not been fully explored. Determining whether these manipulations[2,6] acted by altering a conventional synaptic process (for example, amacrine-cell activity) or levels of protein expression (for example, temporal expression of adhesion molecules) awaits further investigation.

Kerschensteiner and colleagues' findings in the retina[3] contrast with findings in other parts of the nervous system[1,10,11] in which synaptic transmission is essential for determining neuronal structure. During the development of the neuromuscular junction, muscles are initially in contact with multiple motor neurons. But activity-dependent synaptic refinement leads to pruning of axons so that, at maturity, a single axon innervates the muscle[10]. Furthermore, the dendritic structure of frog tectal midbrain neurons, which receive input from retinal ganglion-cell axons, is strongly dependent on the activity-dependent formation of glutamate synapses[1], as is the dendritic structure of Purkinje neurons in the mammalian cerebellum[11]. It will be interesting to learn whether certain features of a neural circuit — such as the level of plasticity at maturity — determine the relative importance of activity-dependent versus molecular-guidance cues for circuit assembly during development. ∎

Jonathan B. Demb is in the Department of Ophthalmology and Visual Sciences, and the Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48105, USA.
Marla B. Feller is in the Department of Molecular and Cell Biology, and the Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA.
e-mails: jdemb@umich.edu;
mfeller@berkeley.edu

1. Cline, H. & Haas, K. *J. Physiol. (Lond.)* **586,** 1509–1517 (2008).
2. Tian, N. *J. Physiol. (Lond.)* **586,** 4347–4355 (2008).
3. Kerschensteiner, D. *et al. Nature* **460,** 1016–1020 (2009).
4. Wässle, H. *Nature Rev. Neurosci.* **5,** 747–757 (2004).
5. Schiavo, G. *et al. Nature* **359,** 832–835 (1992).
6. Bodnarenko, S. R. & Chalupa, L. M. *Nature* **364,** 144–146 (1993).
7. Blankenship, A. G. *et al. Neuron* **62,** 230–241 (2009).
8. Nevin, L. M., Taylor, M. R. & Baier, H. *Neural Dev.* **3,** 36 (2008).
9. Sanes, J. R. & Yamagata, M. *Annu. Rev. Cell Dev. Biol.* doi:10.1146/annurev.cellbio.24.110707.175402 (2009).
10. Sanes, J. R. & Lichtman, J. W. *Annu. Rev. Neurosci.* **22,** 389–442 (1999).
11. Bosman, L. W. & Konnerth, A. *Neuroscience* **162,** 612–623 (2009).

Archaeology of the asteroid belt

John Chambers

The size of asteroids in the Solar System's main asteroid belt may help constrain one of the least-understood aspects of planet formation — the transition from pebble-sized dust balls to mountain-sized planetesimals.

As residents of our own planet Earth, we can all take an interest in how planets are born. The discovery of planets orbiting stars other than the Sun, and the rich variety of worlds that have come to light over the past two decades, make this question even more fascinating. Unfortunately, our understanding of how planets form is frustratingly incomplete, not least because it is extremely difficult to observe this process in action. Writing in *Icarus*, Morbidelli *et al.*[1] shed light on the matter by showing that the early stages of planet formation in the Solar System have left a permanent imprint on the size distribution of objects in the main asteroid belt, which lies between the orbits of Mars and Jupiter.

Planet formation is thought to go through a series of growth stages, beginning with micrometre-sized dust grains that are seen in large numbers in protoplanetary (planet-forming) disks of gas and dust orbiting young stars. Laboratory experiments show that gentle collisions between dust grains cause them to stick together through electrostatic forces, forming millimetre- to centimetre-sized aggregates[2]. Objects that grow larger than about one kilometre — traditionally called planetesimals — have appreciable gravitational fields that hold them together and help them acquire additional material. Low-velocity collisions between planetesimals lead to the formation of Moon-to-Mars-sized 'planetary embryos', and ultimately planets, in a process that is reasonably well understood. However, the transition from pebble-sized dust aggregates to mountain-sized planetesimals is problematic and remains an unresolved issue. This is unfortunate, because all subsequent stages of planet formation depend on it.

Protoplanetary disks are predominantly composed of gas that can frustrate the growth of centimetre- to metre-sized objects in two ways. Gas pressure typically decreases with distance from the central star, and this pressure gradient means that gas orbits the star more slowly than does a solid body. As a result, solid objects experience a headwind, which causes them to drift towards the star. Drift rates increase with size and are greatest for metre-sized bodies, which take as little as 100 years to end up close to the star[3], where they quickly evaporate. This short lifetime of metre-sized bodies means there is a very limited window of opportunity for them to grow into planetesimals, for which radial drift is no longer a factor. Worse still, gas in a protoplanetary disk is probably turbulent. Centimetre-to-metre-sized bodies are strongly stirred by turbulent eddies, so they collide with one another at high speeds, causing them to break apart rather than stick together[4]. For these reasons, it seems unlikely that objects will grow larger than about one metre as a result of the gradual accumulation of dust grains.

Recently, two models have been developed[5,6] that may overcome this 'metre-size barrier'. Each enlists turbulence as an aid to growth rather than a hindrance. The details of the two models differ, but in both cases turbulence acts to concentrate sub-metre-sized particles into diffuse, gravitationally bound clumps 10^4–10^6 km in diameter. These clumps subsequently shrink owing to their own gravity, forming solid objects 100–1,000 km in size, thereby leaping across the metre-size barrier in a single bound. Significantly, both of the models predict that small planetesimals, in the 1–10-km size range, should be rare or even absent.

The new models[5,6] for planetesimal formation are still in their infancy. Ideally, one would like observational data to test their viability. This is where the asteroids come in, because many of these objects are thought to be left over from the time planetesimals and planetary embryos formed. The main asteroid belt has probably gone through four phases of evolution in its history (Fig. 1, overleaf). In stages 1 and 2, planetesimals, and then planetary embryos, formed in the belt as they did elsewhere in the Solar System. In stage 3, the asteroid belt lost more than 99.9% of its mass. Morbidelli *et al.*[1] argue that this material must have been removed dynamically rather than collisionally, probably owing to gravitational perturbations from Jupiter. As a result, the shape of the planetesimal size distribution was preserved during stage 3. The same dynamic process increased the velocities of the planetesimals relative to one another to the high values seen in the asteroid belt today[7]. Stage 4, encompassing most of the history of the Solar System, involved occasional high-speed collisions between planetesimals, causing catastrophic disruptions, and leading to the distribution of the objects we see in the modern asteroid belt.

Previous work[8] has shown that most asteroids larger than 100 km in diameter have survived more or less intact since the end of stage 3, whereas smaller asteroids are likely to be fragments formed in catastrophic collisions between larger bodies. Thus, it seems that, before stage 4, most of the mass in the asteroid

Figure 1 | Evolutionary stages of the main asteroid belt. 1, Dust grains coalesced into planetesimals, objects of 1–1,000 km in diameter, through an unknown process; **2**, planetesimals merged to form Moon-to-Mars-sized planetary embryos; **3**, the vast majority of planetesimals and all embryos were dynamically removed from the asteroid belt; **4**, the surviving planetesimals underwent occasional high-speed collisions with one another, leading to catastrophic break-up and the formation of many small objects. By studying stages **2–4**, and comparing the outcome with the size distribution of present-day asteroids in the main asteroid belt, Morbidelli *et al.*[1] have determined the sizes of planetesimals at the end of stage **1**, placing strong constraints on models of how planetesimals form.

belt existed in bodies larger than 100 km in diameter, and that smaller bodies were relatively rare. Because the dynamic depletion of stage 3 was independent of size, bodies larger than 100 km in diameter also dominated the asteroid belt at the end of stage 2.

Morbidelli and colleagues' contribution[1] is to provide a detailed model of stage 2 (planetesimals to planetary embryos). They consider a wide variety of possible planetesimal size distributions, on the basis that we don't know enough to say how big planetesimals were when they formed. Combining the results of this model with the size-independent depletion of stage 3, the authors find that the current size distribution of the asteroid belt can be reproduced only if most planetesimals were at least 100 km in size at the end of stage 1, with a size distribution extending up to 1,000 km (the size of the belt's largest object, Ceres), and possibly beyond. This is precisely the size range predicted by the new models[5,6] for planetesimal formation. Furthermore, the new study[1] suggests that planetesimals formed with a size distribution that is essentially identical to that seen in the asteroid belt today for diameters larger than 100 km.

The rarity of planetesimals smaller than 100 km in diameter at the end of stage 1 seems to rule out the possibility that dust aggregates somehow made it across the metre-size barrier by gradually sweeping up material from their surroundings. Instead, objects must have grown very rapidly from sub-metre-sized pebbles into 100-km-sized bodies, possibly in a single leap.

Morbidelli and colleagues' model is not the last word on the subject of planet formation. In particular, the authors considered the evolution of the asteroid belt in isolation, neglecting processes that may have exchanged material between the asteroid belt and other parts of the Solar System. However, by showing that the size distribution of the asteroids preserves information about the earliest history of the Solar System, the authors have opened a new window on the least-understood aspect of planet formation. This gives us hope that we may yet understand how Earth and its cousins came to be. ■

John Chambers is in the Department of Terrestrial Magnetism, Carnegie Institution for Science, 5241 Broad Branch Road NW, Washington, DC 20015, USA.
e-mail: chambers@dtm.ciw.edu

1. Morbidelli, A., Bottke, W. F., Nesvorný, D. & Levison, H. F. *Icarus* doi:10.1016/j.icarus.2009.07.011 (2009).
2. Blum, J. & Wurm, G. *Annu. Rev. Astron. Astrophys.* **46,** 21–56 (2008).
3. Weidenschilling, S. J. *Mon. Not. R. Astron. Soc.* **180,** 57–70 (1977).
4. Brauer, F., Dullemond, C. P. & Henning, T. *Astron. Astrophys.* **480,** 859–877 (2008).
5. Johansen, A. *et al. Nature* **448,** 1022–1025 (2007).
6. Cuzzi, J. N. *et al. Astrophys. J.* **687,** 1432–1447 (2008).
7. Petit, J.-M., Morbidelli, A. & Chambers, J. *Icarus* **153,** 338–347 (2001).
8. Bottke, W. F. Jr *et al. Icarus* **175,** 111–140 (2005).

ASTROPHYSICS

Gravity ripples chased

Marc Kamionkowski

Discovering gravitational waves would not only validate Einstein's theory of gravitation but also reveal aspects of the Universe's earliest moments. The hunt for these elusive ripples is now well under way.

For several decades, physicists have been racing to detect gravitational waves — tiny disturbances in space-time that propagate at the speed of light. But the experiments are challenging. They require detection of minute changes — a mere fraction of the size of an atomic nucleus — in kilometre-scale separations between free-floating masses. Although gravitational waves have yet to be detected, on page 990 of this issue the LIGO Scientific Collaboration and the Virgo Collaboration report a milestone in this quest[1]. The authors improve on a long-standing upper limit to the number of gravitational waves generated in the early Universe.

The immediate triumph of Einstein's theory of general relativity was an explanation for the apparently anomalous advance of Mercury's perihelion (its closest orbital point to the Sun) with each orbit. But his theory of gravitation made several other predictions, including the ideas that light from a distant object is deflected by the Sun (confirmed by Arthur Eddington in 1919) and that clocks tick at different rates in different gravitational fields (confirmed by Robert Pound and Glen Rebka in 1959).

Gravitational waves are yet another prediction. They are the gravitational analogues of electromagnetic waves: just as accelerated charges generate electromagnetic waves, accelerated masses produce gravitational waves. Strong, albeit indirect, evidence for the existence of gravitational waves is provided by observations of the Hulse–Taylor binary pulsar, a pair of neutron stars, one of which is a pulsar, revolving around their common centre of mass. The pair's orbital energy loss agrees precisely with that expected in general relativity from the emission of gravitational-wave radiation as the stars draw closer together.

However, it is important to detect gravitational waves directly. That would test whether

they propagate at the speed of light and whether their polarization states agree with those predicted by general relativity. Moreover, because gravitational waves carry information about the sources that produce them and can propagate through regions of the Universe that are opaque to electromagnetic radiation, they offer an otherwise unattainable view of the Universe. For example, whereas the Universe was opaque to electromagnetic radiation during its first 380,000 years, gravitational waves have been able to propagate freely throughout the history of the Universe. The most important pay-off of gravitational-wave experiments will probably be an improved understanding of black holes. The efficiency of gravitational-wave production, relative to that of electromagnetic waves, increases markedly as the sources become denser. Gravitational waves would thus provide much more telling clues about the physics of black holes (the densest objects in the Universe) than do electromagnetic waves.

The investigation undertaken by the LIGO and Virgo collaborations is about a stochastic background signal — originating from all directions in space — of gravitational waves that may have arisen from a large number of unresolved sources and a variety of mechanisms in the early Universe. Such mechanisms include: inflation, a period of exponential space-time expansion that may have occurred as early as 10^{-38} seconds after the Big Bang; early-epoch phase transitions such as that which occurred when electroweak interactions were first broken into distinct electromagnetic and weak interactions about 10^{-11} seconds after the Big Bang; a network of cosmic strings (or other topological defects), a remnant of a phase transition that may have happened before the electroweak one; the 'compactification' of extra spatial dimensions; or some mechanisms associated with string-theory-inspired hypotheses for the origin of the cosmic expansion.

Just as electromagnetic waves are detected through the motions they induce in test charges (free-floating electric charges), gravitational waves can be detected by the motions they induce in test masses. The 'strain' (amplitude) of a gravitational wave determines the fractional change that it induces in the separation between two test masses. However, these motions are expected to be incredibly small. The current generation of experiments is targeting dimensionless strains no larger than 10^{-22}, meaning that measurements must be made of changes that are just a tiny fraction of the size of an atomic nucleus in kilometre-scale distances between objects.

Suppose some mechanism produced gravitational waves before the time of Big-Bang nucleosynthesis (BBN) when protons and neutrons first assembled into helium — a few seconds to minutes after the Big Bang. The energy density of such gravitational waves would have increased the cosmic expansion rate, leaving less time for neutrons to decay, and so more neutrons available to produce

The LIGO site in Livingston, Louisiana.

helium. The observed abundance of helium constrains the fractional gravitational-wave contribution, Ω_{GW}, to the total energy budget of the Universe to less than 1.1×10^{-5} in the frequency band (about 100 Hz) accessible to the LIGO (Laser Interferometer Gravitational-wave Observatory) detector[2,3]. A similar, more recent, bound was obtained from measurements of the cosmic microwave background (CMB) radiation (relic radiation from the hot Big Bang), which can be used to determine the expansion rate of the Universe 380,000 years after the Big Bang[4].

Through an experimental tour de force, the LIGO and Virgo collaborations find that the gravitational-wave background contributes no more than a fraction, $\Omega_{GW} < 7 \times 10^{-6}$, at frequencies near 100 Hz, an order-of-magnitude improvement on their previous result and the aforementioned BBN and CMB bounds. To make these measurements, LIGO uses laser interferometry to determine the distances between test masses (mirror-covered 11-kilogram chunks of fused silica) separated by a distance of 4 km. These measurements constrain the root-mean-square values of the motion of their test masses to be less than a few times 10^{-17}, corresponding to a dimensionless strain of less than 4×10^{-23} over the frequency range 40–170 Hz (see the green curve in Fig. 1 on page 990).

Although the new constraint[1] does not controvert any dearly held theoretical predictions, it represents a watershed event in gravitational-wave astrophysics; terrestrial detectors can now dig into gravitational-wave amplitudes not yet limited by the existing BBN and CMB constraints. And whereas the 'Initial LIGO' detector and Virgo were pursued largely as engineering projects (there were no really promising gravitational-wave sources accessible to Initial LIGO and Virgo), the next generation of detectors, Advanced LIGO[5] (due to begin observing in 2014) and Advanced Virgo[6] are almost guaranteed to see

a signal from compact binary-star systems.

But there are other, albeit indirect, gravitational-wave detectors that offer competition to LIGO and Virgo. Pulsar-timing arrays[7], which measure tiny gravitational-wave-induced shifts in the arrival times of radio pulses from a collection of pulsars, are racing to achieve indirect detection of gravitational waves from merging supermassive black holes. Experiments that measure the CMB, such as the recently launched Planck satellite[8], will seek to detect ultralow-frequency gravitational waves from inflation through their imprint on the polarization of CMB radiation.

By exploiting the effect of higher-frequency gravitational waves on the temperature of CMB radiation[4], CMB experiments can also improve current sensitivity to such waves by an order of magnitude. These indirect measurements may come in before Advanced LIGO/Virgo start operation. But dedicated gravitational-wave observatories, which may later include the spaced-based Laser Interferometer Space Antenna (LISA)[9], will be required to capitalize fully on the new observational window that gravitational waves will ultimately provide. LISA will operate at the lower frequencies at which the signals from compact stellar binaries and supermassive black holes are expected to be even clearer than in the LIGO/Virgo frequency band. ∎

Marc Kamionkowski is at the California Institute of Technology, Pasadena, California 91125, USA.
e-mail: kamion@tapir.caltech.edu

1. The LIGO Scientific Collaboration and the Virgo Collaboration Nature 460, 990–994 (2009).
2. Allen, B. in Proceedings of the Les Houches School on Astrophysical Sources of Gravitational Waves (eds Marck, J.-A. & Lasota, J.-P.) 373–417 (Springer, 1996).
3. www.ligo.caltech.edu
4. Smith, T. L., Pierpaoli, E. & Kamionkowski, M. Phys. Rev. Lett. 97, 021301 (2006).
5. www.ligo.caltech.edu/advLIGO/scripts/summary.shtml
6. wwwcascina.virgo.infn.it/advirgo
7. Demorest, P., Lazio, J. & Lommen, A. Preprint at http://arxiv.org/abs/0902.2968 (2009).
8. http://planck.esa.int
9. www.lisa.nasa.gov

HYPOTHESIS

Evidence for an early prokaryotic endosymbiosis

James A. Lake[1,2,3,4]

Endosymbioses have dramatically altered eukaryotic life, but are thought to have negligibly affected prokaryotic evolution. Here, by analysing the flows of protein families, I present evidence that the double-membrane, Gram-negative prokaryotes were formed as the result of a symbiosis between an ancient actinobacterium and an ancient clostridium. The resulting taxon has been extraordinarily successful, and has profoundly altered the evolution of life by providing endosymbionts necessary for the emergence of eukaryotes and by generating Earth's oxygen atmosphere. Their double-membrane architecture and the observed genome flows into them suggest a common evolutionary mechanism for their origin: an endosymbiosis between a clostridium and actinobacterium.

Endosymbioses have affected almost every aspect of eukaryotic evolution[1], but appreciation of the pervasiveness of this mechanism has not been achieved without controversy. Virtually every proposed symbiotic event in the origin of the eukaryotic cell has been vigorously debated[2–7]. Nevertheless, today there is general agreement that the chloroplast and mitochondria have endosymbiotic origins, and there is a developing recognition that eukaryotic cells have hosted diverse endosymbiotic guests during their existence. These include numerous and varied prokaryotes, and even other eukaryotes complete with nucleus and organelles.

Acting as functional modules, endosymbionts have the ability to dramatically transform the metabolic and architectural properties of their hosts[8]. Endosymbiotic guests bring with them novel metabolic and signalling capabilities. These new capabilities can be provided through the introduction of new genes, or through the generation of additional membrane-bounded compartments formed during endosymbioses. The ability to form new cellular compartments distinguishes endosymbioses from repeated episodes of gene transfer involving conjugation, transformation and transduction[8–10].

These are exciting times for microbiology, with important discoveries being made almost every week[11–13]. Here I explore what is known about prokaryotic symbioses and endosymbioses, and provide evidence that a significant fraction of prokaryotic diversity has been affected by symbioses, and probably by endosymbioses.

Exploring the possibility of prokaryotic endosymbioses

It is widely accepted that eukaryotes have frequently hosted endosymbionts during the course of their evolution. But can prokaryotes themselves also host other endosymbiotic guests? Apparently so, at least within eukaryotic cells. There is good evidence for a eukaryote containing a prokaryotic host and guest in the secondary endosymbionts of certain aphids[14]. Specifically, the β-proteobacterial *Buchnera* endosymbionts themselves contain γ-proteobacterial endosymbionts, generating endosymbionts within other endosymbionts, much like a set of Russian nesting (*matryoshka*) dolls. Resembling eukaryotic endosymbioses, the γ-proteobacterial endosymbionts are present within the cytoplasm of their β-proteobacterial hosts (Fig. 1a).

Exclusively prokaryotic endosymbioses are unknown, although the phototrophic consortium 'Chlorochromatium aggregatum' comes pretty close. It consists of two partners; a motile, heterotrophic, central β-proteobacterium surrounded by 10–60 peripheral photosynthetic green sulphur bacteria[15]. The green sulphur bacteria, or epibionts, labelled GSB in Fig. 1b, are attached to the central bacterium through periplasmic tubules formed by extensions from the outer membrane of the flagellated central β-proteobacterium. This allows the central β-proteobacterium, labelled β in Fig. 1b, to move the consortium to favourable anaerobic habitats where sulphides and sufficient light for photosynthesis are available. Structural studies of the consortium suggest that, ''the two partner bacteria may actually share a common periplasmic space''[16]. If so, then both partners would be enclosed by a common periplasm, but not by a common membrane. Although this symbiosis does not provide a precedent for a prokaryotic endosymbiosis, it suggests how one might have started.

Could it be possible that primary endosymbioses exclusively involving prokaryotes have shaped the evolution of life at some time in the past?

Prokaryotic diversity

There is currently much discussion of the prokaryotic 'tree of life', but there are few points of agreement regarding its topology, except
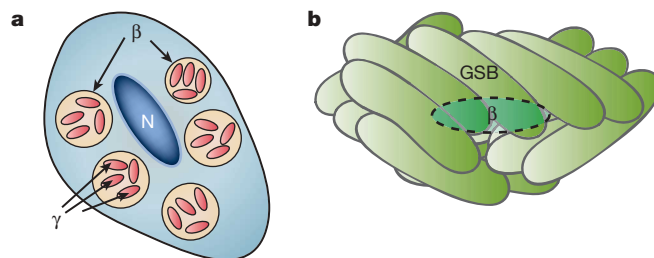
Figure 1 | Schematic diagrams illustrating prokaryotic symbionts. a, A diagram of a mealybug cell containing β-proteobacterial endosymbionts that contain γ-proteobacterial endosymbionts within them[14]. The mealybug nucleus, N, is surrounded by β-proteobacterial endosymbionts, labelled β, that are themselves hosts to γ-proteobacterial endosymbionts, labelled γ. **b**, A diagram of the prokaryotic consortium, *Chlorochromatium aggregatum*[16], consisting of peripheral green sulphur bacteria, labelled GSB, surrounding a central, spindle shaped, β-proteobacterium, shown by a dashed line and labelled β.

[1]Department of Molecular, Cellular and Developmental Biology, [2]Molecular Biology Institute, [3]Department of Human Genetics, [4]UCLA Astrobiology Institute, University of California, Los Angeles, California 90095, USA.

that it is not a tree[17–25]. However, if we blur our taxonomic vision and just look at the low-resolution outlines of life, then it is generally agreed that all prokaryotes belong to one of five groups[26–29]. These five natural, phylogenetically well separated, and relatively homogeneous taxa[30–32] are: the Archaea, the Bacilli and relatives, the Clostridia and relatives, the Actinobacteria, and the double membrane, or Gram negative, prokaryotes. The Archaea contain extreme halophiles, methanogens, hyperthermophiles and other unique phenotypes. The Bacilli and the Clostridia are firmicutes, characterized by their low-guanine-cytosine (low-GC) genomic compositions, although not exclusively[33]. The Clostridia are unique among the single-membrane prokaryotes for including photosynthetic organisms as well as fermenting ones. The Actinobacteria, characterized by high-GC genomic compositions, are morphologically diverse and include many human pathogens, such as those that cause leprosy and tuberculosis. The double-membrane, Gram-negative, prokaryotes encompass approximately 42 phyla. (For a detailed list of the phyla, including some recent taxonomic changes[26–28], see Supplementary Information section 1.) Together, these five super-taxa contain all known prokaryotic life[34].

The double-membrane prokaryotes

Consider the double-membrane prokaryotes in a little more detail. This taxon contains an amazingly diverse, inordinately speciose, and possibly primitively photosynthetic group of organisms. It encompasses more proteomic diversity as measured by protein families (4,756) than all other prokaryotes combined (4,000). Furthermore, they have profoundly altered life on Earth, producing 20% of the Earth's atmosphere and providing endosymbionts critical for the emergence of eukaryotes. It includes the photosynthetic Cyanobacteria, α-, β-, γ-, δ- and ε-proteobacteria, Chloroflexi and Chlorobi, and numerous other intriguing taxa, such as the Spirochaetes, Planctomycetes and Aquificales.

Before the advent of molecular sequencing, the double-membrane clan, an unrooted clade[35], was loosely distinguished from single-membrane prokaryotic groups by its Gram-negative membrane staining properties[36]. Subsequently, its membership was defined more accurately when phyla within the double-membrane clan were identified on the basis of the presence of characteristic indels (that is, insertions or deletions) contained within proteins[37]. Most recently, indel rooting studies have shown that the double-membrane clan is a clade, based on heat-shock protein Hsp70 (ref. 31) and on enzymes involved in pyrimidine-biosynthesis (PyrD protein) and histidine-biosynthesis (HisA and HisF proteins)[29]. It is interesting that these last two indels also showed that the double-membrane prokaryotes are closely related to the Actinobacteria. This point will be discussed further below.

It is also intriguing that they have this double-membrane structure, whereas other prokaryotic groups are surrounded by single membranes. In fact, the membrane organization of double-membrane prokaryotes fundamentally differs from that found in single-membrane prokaryotes. In the former, the peptidoglycan layer is sandwiched between the outer and inner membranes, so that it surrounds the inner membrane: in contrast, in the latter there is no inner membrane, and the peptidoglycan layer, located outside the cell, surrounds the outer membrane. Also, double-membrane prokaryotes contain their flagellar motors in the inner membrane, whereas single-membrane prokaryotes contain their flagellar motors in the outer membrane. And the photosynthetic apparatus in double-membrane prokaryotes is in the inner membrane, rather than in the outer membranes as in single-membrane prokaryotes. In other words, the organization of the inner membrane of the double-membrane prokaryotes resembles that of the outer membranes of typical single-membrane prokaryotes. The inner membranes of double-membrane prokaryotes are organized almost as if they were derived from the outer membrane of an engulfed single-membrane prokaryote.

Phylogenetic testing

If the double-membrane prokaryotes were produced by an endosymbiosis, then perhaps the genomic signatures of the endosymbiosis might still be strong enough to be detected in phylogenetic reconstructions. Here I reconstruct the evolution of the double-membrane prokaryotes using methods capable of discriminating between rings, caused by endosymbioses, and trees, caused by clonal evolution. These methods[38,39] distinguish trees from rings by analysing gene, or protein family, presences and absences. In many ways they are similar to, but in other ways quite different from, tree reconstruction algorithms that use nucleotide and amino acid changes.

If the double-membrane prokaryotes were the result of an endosymbiosis, then the flow of genes from the two donors into the proto-double-membrane-prokaryote would not be consistent with any tree, as branches on a tree must diverge from a single node representing a common ancestral organism. In contrast, an endosymbiosis, representing the genomic merger of two distinct organisms into one, proceeds from two separate nodes corresponding to the host and the guest. Thus endosymbioses produce rings rather than trees, as in Fig. 2a, similar to that observed for the eukaryotic ring of life[20]. If the present studies were to reconstruct a tree rather than a ring, then an endosymbiosis would be rejected, thereby providing a first test for an endosymbiosis.

But if the double membranes are signs of a past endosymbiosis, then which prokaryotes might be the culprits—that is, the host and the guest? Fortunately, because genes can flow into the new organism from two different donors during an endosymbiosis, phylogenetic analyses can also identify the donors. In the case of an endosymbiosis, the two donor taxa surrounding the clade of double-membrane prokaryotes (D) would correspond to the donor taxa (Fig. 2b). Thus if the taxa adjacent to D could be identified, this would provide evidence for specific donors, and represent a second level of testing for an endosymbiosis.

And finally, how could one determine which taxon within the ring would be the endosymbiont? For example, if a ring of five taxa is reconstructed, at first glance it might seem that any of the five could be the endosymbiont. As demonstrated elsewhere[38], presence-absence analyses are able to detect the directions of gene flow. In an endosymbiosis, one expects to see genes flowing from both donors into the endosymbiont (Fig. 2c). Thus if genes were found to flow into the double-membrane prokaryotes from both sides, then this pattern of gene flow, and only this pattern, would be consistent with a double-membrane endosymbiotic origin. Accordingly, gene presence-absence studies can specifically identify the endosymbiont and thereby provide a third level of proof for an endosymbiosis.
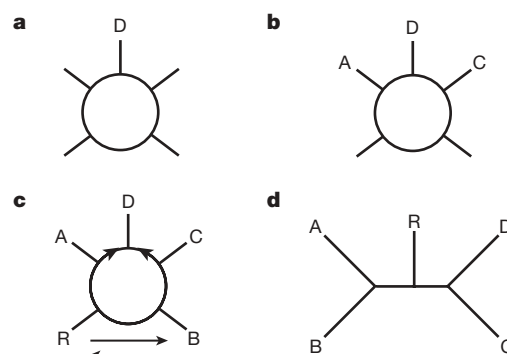
Figure 2 | An illustration of three steps required for the identification of an endosymbiont from gene flow data, and the tree of life that best fits the gene flow data. a, The endosymbiont, D, must be part of a ring. **b**, The genome donors, A and C, must surround the endosymbiont and be fully resolved. **c**, The direction of gene flow must be from both donors to the endosymbiont, D. **d**, The best tree of life. A, Actinobacteria; B, Bacilli; C, Clostridia; D, Double-membrane prokaryotes; R, Archaea.

However, this still might not distinguish an endosymbiosis from alternative types of symbioses, but that point will be discussed subsequently.

Protein family analyses

To test for prokaryotic endosymbioses, I analyse here the five natural taxa previously described: the Archaea (R), the Actinobacteria (A), the Bacilli and relatives (B), the Clostridia and relatives (C), and the double-membrane prokaryotes (D). Using absence-presence analyses of protein families obtained from >3,000 diverse prokaryotes, I reconstruct a low-resolution graph of ancient prokaryotic evolution, to determine whether it is a tree, a ring, or some other graph.

The information required to reconstruct the 'tree/graph of life' is contained in ten parsimoniously informative patterns of protein family absences and presences[38,39]. These patterns are listed in the first five columns of Table 1, and consist of three gene presences ('+'), and two gene absences (blanks). In the first row, three protein families present in Archaea (R), Actinobacteria (A) and Bacilli (B), and absent from all Clostridia (C) and all double-membrane prokaryotes (D) are analysed. The most parsimonious tree or graph is determined by computing the character state patterns allowed by graphical models, as described elsewhere[38], and then comparing these 'allowed patterns' with the observed numbers of protein families that support each pattern.

As it might happen that the data have little resolving power and are unable to discriminate readily between tree-like evolution and endosymbiotic evolution, I explicitly compare the graph that best fits the data with the tree that best fits the data to see if trees can be resolved from rings. The number of protein families observed for each pattern is shown in the sixth and seventh columns of Table 1. Bold font with asterisk indicates that large numbers of counts are predicted, and standard fonts indicate that small numbers of counts are predicted[38]. Predictions based on the best graph (Fig. 2c) are shown in the sixth column, and those based on the best tree (Fig. 2d) are shown in the seventh column.

It is immediately clear that the data fit the predictions calculated from the best graph. The five allowed patterns coincide with the five largest scores, 62–174, and the five disallowed patterns coincide with the five smallest scores, 0–15. But might they, through some statistical quirk, also fit the best tree? For the best tree, the data conflict with the predictions for three patterns. The conflicts are: large scores are predicted by the best tree for the patterns in the first and fourth rows and yet they correspond to small counts, 3 and 15; and small

scores are predicted for the pattern in row five and yet it corresponds to a large count, 62. Clearly, the significantly better agreement found for the best graph compared with the best tree indicates that graphs are needed to explain prokaryotic evolution.

Furthermore, there is unusually strong statistical support for the best graph. Using the numbers of allowed protein families to estimate the probabilities of each of the ten informative character state sets provides significant posterior bootstrap support[38,40] for the graph in Fig. 2c, $P > 0.9999$, whereas there is no support for the tree in Fig. 2d, $P < 0.0001$.

Indeed, when the evidence based on the flow of protein families is examined, it is certainly plausible that it was an endosymbiosis. In Fig. 3, the taxa meet all of the criteria supporting an endosymbiosis. (1) The putative endosymbiont, the double-membrane taxon, is part of the ring. (2) The double-membrane prokaryotes are surrounded on the left by the Actinobacteria, and on the right by the Clostridia, indicating that these two taxa are the donors. (3) The arrows on either side of the double-membrane prokaryotes indicate that genes flow from the Actinobacteria and from the Clostridia, implicating them as the gene donors, and flow into the double-membrane prokaryotes, specifically identifying them as the recipient.

The topology of the graph in Fig. 3 is supported by some recent, formerly perplexing observations. Analyses of the distribution of two slowly evolving indels found in enzymes involved in pyrimidine- and histidine-biosynthesis, PyrD and HisA, demonstrated that the double-membrane prokaryotes and the Actinobacteria form a statistically supported clade[29]. This supports the upper-left portion of the prokaryotic ring of life, where genes flow from the Actinobacteria into the double-membrane prokaryotes. Furthermore, as the PyrD and HisA indels were analysed with respect to paralogous outgroups, they established that genes flowed from the root into the Actinobacteria and the double-membrane prokaryotes, consistent with the direction of gene flow shown in Fig. 3. Thus these indel analyses provide statistically significant, independent support for the left half of the symbiosis.

The graph in Fig. 3 also starts to make sense of the perplexing phylogenetic distribution of the photosynthetic machinery. Complex, highly integrated systems involving many genes are thought to be more difficult to transfer horizontally/laterally[10,41]. As the photosynthetic apparatus involves scores of genes, this makes the presence or absence of photosynthesis a reasonable phylogenetic marker for the study of ancient prokaryotic divergences. In prokaryotes, photosynthesis is found only in the double-membrane prokaryotes and in the Clostridia[42]. This distribution has been hard to reconcile with trees. But it now starts to fit with Fig. 3, because at the upper right-hand side of the figure the double-membrane prokaryotes are adjacent to the Clostridia. Thus, consistent with photosynthesis originating in the Clostridia, this graph parsimoniously explains the transfer of

Table 1 | Protein family support for the best ring and for the best tree of prokaryotic life

Character state patterns					Best graph protein families	Best tree protein families
R	A	B	C	D		
+	+	+			3	**3***
+	+		+		0	0
+		+	+		8	8
	+	+	+		15	**15***
+	+			+	**62***	62
+		+		+	15	15
	+	+		+	**91***	**91***
+			+	+	**99***	**99***
	+		+	+	**73***	**73***
		+	+	+	**174***	**174***

The topology corresponding to the 'best ring' of prokaryotic life is shown in Fig. 2c and that corresponding to the 'best tree' is shown in Fig. 2d. Protein families are calculated using the Pfam (v. 22.0) database as implemented on the Janelia Farms website (available at http://pfam.janelia.org/search). Allowed character state patterns are indicated in the table by bold type and are calculated according to procedures published elsewhere[38,39]. Specifically, the five allowed patterns predicted by the best ring of life, shown in Fig. 2c, are those in which A and D are both '+', and/or those in which C and D are both '+'. The six allowed patterns predicted by the best tree of life, shown in Fig. 2d, are those in which A and B are both '+' and/or those in which C and D are both '+'. Explicit definitions of the prokaryotic groups included within the Archaea, R, the Actinobacteria, A, the Bacilli and relatives, B, the Clostridia and relatives, C, and the double-membrane prokaryotes, D, are provided in Supplementary Table 1. Detailed listings of the protein families that are taxonomically distributed according to each of the ten character state patterns are provided in Supplementary Information section 2.
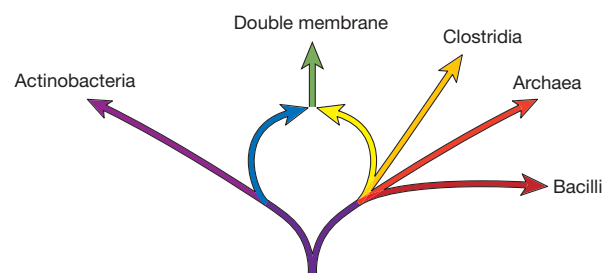
Figure 3 | A schematic diagram illustrating the prokaryotic ring of life. The actinobacterial genome donor, at the left (blue), and the clostridial genome donor, at the right (yellow), transfer their genomes to form the double-membrane prokaryotes at the top of the ring (green). The protein family data identify the Actinobacteria and the Clostridia as donors, and the double-membrane prokaryotes as the fusion organism, but cannot fully resolve the relationship between the Bacilli (dark red) and the Archaea (bright red)[38]. This ring partially explains why a prokaryotic 'tree of life' is not a tree[23]. The rooting shown is the new root of life[29,32,51].

969

photosynthetic genes to the double-membrane prokaryotes. This, of course, does not resolve the many questions related to the evolution of photosystems I and II[43], but it is a start.

These results are certainly consistent with the proposition that the double-membrane prokaryotes were formed as the result of an endosymbiosis between an ancient ancestral population of Clostridia and Actinobacteria. So far I have assumed that an endosymbiosis was responsible, but might other symbioses also be possible?

Symbioses and genome transfers

Symbioses, including endosymbioses, do not happen in a single generation. They develop over long periods of time as symbiotic partners evolve, adapt, and exchange genes through the traditional mechanisms of gene transfer—that is, conjugation, transformation, and viral transduction[9,10]. During a symbiosis, frequent gene transfers occur between partners owing to physical proximity and other factors[10,19,44–46]. Thus it is thought that, over time, organisms in stable symbioses can transfer significant portions of their genomes to their partners, resulting in the nearly complete transfer of genomes, or 'genome transfer'[47]. On the basis of the protein family results and analyses presented here, it is also possible that various types of symbioses could have produced the protein family results in Table 1, provided that these symbioses are also consistent with those analyses. Thus any two-donor symbioses involving a clostridium and an actinobacterium donating their genes to form the double-membrane prokaryotes would fit these data. This very probably happened over an extended period of time.

We cannot say exactly how much time this symbiosis required, but it definitely did not happen in the past two billion years. We know this because the cyanobacterial double-membrane prokaryotes are responsible for producing the Earth's oxygen atmosphere. This implies that their diversification, and hence that of the double-membrane clade, started before the rise of oxygen in the atmosphere, approximately 2.4 billion years ago[48,49] or possibly earlier (2.7 billion years ago[48,50]). Because the double-membrane prokaryotes are descended from Actinobacteria and Clostridia, these two donor clades must trace their beginnings back to even earlier times. Thus there seem to be no obvious physical time constraints on how long the initial phase of the double-membrane symbiosis lasted. In the future, the double-membrane clade may provide a useful reference taxon for calibrating molecular clock studies.

Symbiosis or endosymbiosis

I find it fascinating that this prokaryotic symbiosis could so profoundly shape the evolution of life, and thereby set the stage for the formation of an oxygen-rich atmosphere and for the emergence of eukaryotic organelles. The symbiosis of two disparate prokaryotes to form the double-membrane prokaryotes appears to have done more than just introduce a new combination of genes into a cell: it also generated a group of organisms surrounded by two membranes. In some ways, it appears as if the formation of the periplasm, the movement of the peptidoglycan layer from outside the cell into the periplasm, the movement of the flagella motor into the inner membrane, and the redesign of older transport proteins into those capable of spanning two membranes represents a radically new structural design.

I cannot help but notice that the existence of the double-membrane structure immediately suggests a possible mechanism for its formation and for the observed genome transfers from Clostridia and Actinobacteria to the double-membrane prokaryotes—endosymbioses. I believe that this agreement should not be ascribed to chance.

1. Margulis, L. *Origin of the Eukaryotic Cells* (Yale Univ. Press, 1970).
2. Margulis, L. *Symbiosis in Cell Evolution* (Freeman and Company, 1981).
3. Lake, J. A. & Rivera, M. C. Was the nucleus the first endosymbiont? *Proc. Natl Acad. Sci. USA* **91**, 2880–2881 (1994).
4. Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
5. Lang, B. F. & Gray, M. W. Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397 (1999).
6. Poole, A., Jeffares, D. & Penny, D. Early evolution: prokaryotes, the new kids on the block. *Bioessays* **21**, 880–889 (1999).
7. Keeling, P. J. & Palmer, J. D. Horizontal gene transfer in eukaryotic evolution. *Nature Rev. Genet.* **9**, 605–614 (2008).
8. Archibald, J. M. & Keeling, P. J. Recycled plastids: a 'green movement' in eukaryotic evolution. *Trends Genet.* **18**, 577–584 (2002).
9. Syvanen, M. & Kado, C. I. (eds) *Horizontal Gene Transfer* (Chapman and Hall, 1998).
10. Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. Horizontal gene transfer in microbial genome evolution. *Theor. Popul. Biol.* **61**, 489–495 (2002).
11. Bowler, C., Karl, D. M. & Colwell, R. R. Microbial oceanography in a sea of opportunity. *Nature* **459**, 180–184 (2009).
12. Fuhrman, J. A. Microbial community structure and its functional implications. *Nature* **459**, 193–199 (2009).
13. DeLong, E. F. The microbial ocean from genomes to biomes. *Nature* **459**, 200–206 (2009).
14. von Dohlen, C. D., Kohler, S., Alsop, S. T. & McManus, W. R. Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature* **412**, 433–436 (2001).
15. Froestl, J. & Overmann, J. Phylogenetic affiliation of the bacteria that constitute phototrophic consortia. *Arch. Microbiol.* **174**, 50–58 (2000).
16. Wanner, G., Vogl, K. & Overmann, J. Ultrastructural characterization of the prokaryotic symbiosis in "*Chlorochromatium aggregatum*". *J. Bacteriol.* **190**, 3721–3730 (2008).
17. Hilario, E. & Gogarten, J. P. Horizontal transfer of ATPase genes—the tree of life becomes a net of life. *Biosystems* **31**, 111–119 (1993).
18. Dagan, T. & Martin, W. The tree of one percent. *Genome Biol.* **7**, 1181–1187 (2006).
19. Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**, 1598–1602 (2003).
20. Rivera, M. C. & Lake, J. A. The ring of life: evidence for a genome fusion origin of eukaryotes. *Nature* **431**, 152–155 (2004).
21. Konstantinidis, K. T. & Tiedje, J. M. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* **187**, 6258–6264 (2005).
22. Sorek, R. *et al.* Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**, 1449–1452 (2007).
23. Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284**, 2124–2128 (1999).
24. McInerney, J. O. & Pisani, D. Paradigm for life. *Science* **318**, 1390–1391 (2007).
25. Martin, W. & Embley, T. M. Early evolution comes full circle. *Nature* **431**, 134–136 (2004).
26. Ohno, M. *et al.* Symbiobacterium thermophilum gen. nov., sp. nov., a symbiotic thermophile that depends on co-culture with a *Bacillus* strain for growth. *Int. J. Syst. Evol. Microbiol.* **50**, 1829–1832 (2000).
27. Garrity, G. & Holt, J. G. in *Bergey's Manual of Systematic Bacteriology* (eds Boone, D. & Castenholz, R. W.) 119–141 (Springer, 2001).
28. Wu, M. *et al.* Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet.* **1**, 563–574 (2005).
29. Lake, J. A., Servin, J. A., Herbold, C. W. & Skophammer, R. G. Evidence for a new root of the tree of life. *Syst. Biol.* **57**, 835–843 (2008).
30. Skophammer, R. G., Herbold, C. W., Rivera, M., Servin, J. A. & Lake, J. A. Evidence that the root of the tree of life is not within the Archaea. *Mol. Biol. Evol.* **23**, 1648–1651 (2006).
31. Lake, J. A., Herbold, C. W., Rivera, M. C., Servin, J. A. & Skophammer, R. G. Rooting the tree of life using non-ubiquitous genes. *Mol. Biol. Evol.* **24**, 130–136 (2007).
32. Skophammer, R. G., Servin, J. A., Herbold, C. W. & Lake, J. A. Evidence for a Gram positive, eubacterial root of the tree of life. *Mol. Biol. Evol.* **24**, 1761–1768 (2007).
33. Ueda, K. *et al.* Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. *Nucleic Acids Res.* **32**, 4937–4944 (2004).
34. Boone, D. & Castenholz, R. W. (eds) *Bergey's Manual of Systematic Bacteriology* 2nd edn Vol. 1, *The Archaea and the Deep Branching and Phototrophic Bacteria* (ed. Garrity, G. M.) (Springer, 2001).
35. Wilkinson, M., McInerney, J. O., Hirt, R. P., Foster, P. G. & Embley, T. M. Of clades and clans: terms for phylogenetic relationships in unrooted trees. *Trends Ecol. Evol.* **22**, 114–115 (2007).
36. Stanier, R. Y., Adelberg, E. A. & Ingraham, J. L. *The Microbial World* (Prentice-Hall, 1976).
37. Gupta, R. S. Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* **62**, 1435–1491 (1998).
38. Lake, J. A. Reconstructing evolutionary graphs: 3D parsimony. *Mol. Biol. Evol.* **25**, 1677–1682 (2008).
39. Lake, J. A. & Rivera, M. C. Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction. *Mol. Biol. Evol.* **21**, 681–690 (2004).
40. Lake, J. A. Calculating the probability of multitaxon evolutionary trees—bootstrappers gambit. *Proc. Natl Acad. Sci. USA* **92**, 9662–9666 (1995).

41. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96,** 3801–3806 (1999).

42. Gest, H. & Favinger, J. L. *Heliobacterium chlorum* gen. nov. sp. nov., an anoxygenic brownish-green photosynthetic bacterium containing a new form of bacteriochlorophyll. *Arch. Microbiol.* **136,** 11–16 (1983).

43. Raymond, J., Zhaxybayeva, O., Gogarten, J. P., Gerdes, S. Y. & Blankenship, R. E. Whole-genome analysis of photosynthetic prokaryotes. *Science* **298,** 1616–1620 (2002).

44. Boone, D. R., Castenholz, R. W., Johnson, R. L. & Liu, Y. Diffusion of the interspecies electron carriers H_2 and formate in methanogenic ecosystems and its implications in the measurement of K_m for H_2 or formate uptake. *Appl. Environ. Microbiol.* **55,** 1735–1741 (1989).

45. Overmann, J. & Schubert, K. Phototropic consortia: model systems for symbiotic interrelations between prokaryotes. *Arch. Microbiol.* **177,** 201–208 (2002).

46. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428,** 37–43 (2004).

47. Margulis, L. in *Genomes in Flux* (eds Gogarten, M. Olendzenski, L. & Gogarten, J. P.) (Humana, in the press).

48. Fischer, W. W. Life before the rise of oxygen. *Nature* **455,** 1051–1052 (2008).

49. Rasmussen, B., Fletcher, I. R., Brocks, J. J. & Kilburn, M. R. Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* **455,** 1101–1104 (2008).

50. Ventura, G. T. *et al.* Molecular evidence of Late Archean archaea and the presence of a subsurface hydrothermal biosphere. *Proc. Natl Acad. Sci. USA* **104,** 14260–14265 (2007).

51. Servin, J. A., Herbold, C. W., Skophammer, R. G. & Lake, J. A. Evidence excluding the root of the tree of life from the Actinobacteria. *Mol. Biol. Evol.* **25,** 1–4 (2008).

ARTICLES

High harmonic interferometry of multi-electron dynamics in molecules

Olga Smirnova[1,2], Yann Mairesse[1,3], Serguei Patchkovskii[1], Nirit Dudovich[1,4], David Villeneuve[1], Paul Corkum[1] & Misha Yu. Ivanov[1,5]

High harmonic emission occurs when an electron, liberated from a molecule by an incident intense laser field, gains energy from the field and recombines with the parent molecular ion. The emission provides a snapshot of the structure and dynamics of the recombining system, encoded in the amplitudes, phases and polarization of the harmonic light. Here we show with CO_2 molecules that high harmonic interferometry can retrieve this structural and dynamic information: by measuring the phases and amplitudes of the harmonic emission, we reveal 'fingerprints' of multiple molecular orbitals participating in the process and decode the underlying attosecond multi-electron dynamics, including the dynamics of electron rearrangement upon ionization. These findings establish high harmonic interferometry as an effective approach to resolving multi-electron dynamics with sub-Ångström spatial resolution arising from the de Broglie wavelength of the recombining electron, and attosecond temporal resolution arising from the timescale of the recombination event.

Structure-induced interference minima in photo-ionization were predicted for H_2^+ by Cohen and Fano[1], providing the foundation for extracting structural information from photoelectron spectra. Similar interference in photo-recombination, which is the last step in high harmonic emission, has been predicted to produce minima in the harmonic spectra of H_2^+ (ref. 2), stimulating experiments on using high harmonic generation to image molecular structures[3–11]. These experiments benefit from the ability to induce and control molecular alignment in a gas[12,13]. The minima have indeed been found in the harmonic spectra of the CO_2 molecule, but at vastly different positions in different experiments[4,5,8,9]. The structure of the molecule is the same, so why does the minimum appear at different places? We find that it shifts with the laser intensity and its position and mobility reflect attosecond multielectron dynamics in the ion, including electron rearrangement upon ionization.

It has been assumed that structures in the harmonic spectra reflect the geometry of the highest occupied molecular orbital (HOMO)[3–5,8,9,14–18]. Indeed, strong-field ionization is exponentially sensitive to the ionization potential I_p, suggesting that after ionization the molecular ion is left in its ground electronic state (electron removal from HOMO in the Hartree–Fock picture). In this case, even though rearrangement has occurred, only structural information will be recorded in the harmonic spectrum. However, in molecular ions electronic excitations often lie within a couple of electron volts from the ground state, leading to sizeable population amplitudes even for exponential scaling of ionization rates. The geometry of the molecular orbitals is also crucial[19,20] and can further increase relative populations of the excited electronic states of the ion (tunnelling from deeper orbitals in the Hartree–Fock picture). The importance of different orbitals in harmonic generation has been pointed out in the experiments[10,11].

Using the example of a CO_2 molecule, we identify participating orbitals by measuring the phase of the high harmonics using two-slit interferometry[21]. In one arm we place randomly aligned molecules (a reference) and in the other, molecules aligned at a specific angle θ.

Combined with amplitude measurements, the phase of the harmonic emission can give a 'fingerprint' of the molecular orbital. We identify at least three orbitals, and three states of the ion spanning a 4.3 eV range of ionization potentials, that contribute to harmonic emission in CO_2 molecules. Varying the laser intensity and controlling molecular alignment relative to laser polarization allows us to vary the contribution of each orbital to a given harmonic.

Access to multiple electronic states of the ion implies the potential for inducing and imaging multi-electron dynamics in the ion. We use high harmonic interferometry to reconstruct the shape and the motion of the hole created in the molecule upon ionization, gaining insight into attosecond dynamics of electron rearrangement during tunnelling.

There are two fundamental steps in resolving multi-electron dynamics: the identification of multiple orbitals (ionic states) participating in the process and the observation of their interference.

Guidelines for tracking multiple orbitals

When an intense laser pulse ionizes a molecule, its wavefunction splits into two components: a free electron wavepacket and its associated ion, and the neutral molecule. The laser field pushes the liberated electron away from the ion and then drives it back[22]. There, the electron can recombine with the ion, re-forming the neutral molecule in its original state and producing coherent radiation—high harmonics of the incident light. Different states of the molecular ion between ionization and recombination provide different 'channels' for this process (Fig. 1).

For each channel, specific electronic structure is encoded onto the harmonic amplitude and phase. These channels are different pathways that take the system from the same initial to the same final state via different states of the ion. They provide a basis for recording multielectron dynamics, including dynamics of electron rearrangement upon ionization. Their interference records the relative phases between the channels by mapping them onto the amplitude modulation in the harmonic spectra. The phases between the channels,

[1]National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada. [2]Max-Born Institute, 2a Max-Born-Strasse, Berlin D-12489, Germany. [3]CELIA, Université Bordeaux I, UMR 5107 (CNRS, Bordeaux 1, CEA), 351 Cours de la Libération, 33405 Talence cedex, France. [4]Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel. [5]Department of Physics, Imperial College London, South Kensington Campus, SW7 2AZ London, UK.
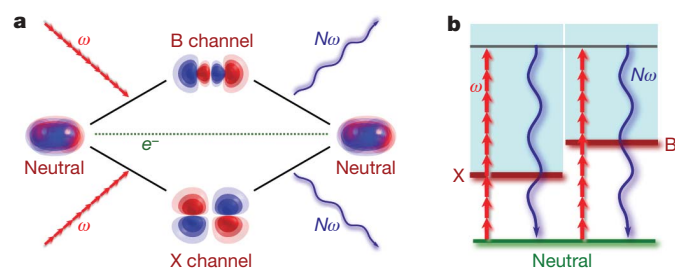
Figure 1 | Pictorial schemes of multichannel interference in high harmonic generation, for two channels. The orbitals sketched in **a** correspond to the two relevant channels X and B in a CO_2 molecule aligned parallel to the laser polarization. Channels X and B are associated with $\tilde{X}\,^2\Pi_g$ and $\tilde{B}\,^2\Sigma^+_u$ states of the cation (HOMO and HOMO − 2 orbitals in the Hartree–Fock picture). Panel **b** illustrates channel interference in terms of a conventional nonlinear optics diagram.

accumulated between ionization and recombination, encode the underlying electronic dynamics with attosecond time resolution. High time resolution arises from the fact that different harmonics are emitted at different time-delays between ionization and recombination, within a fraction of a laser cycle[6,7,23,24]. Each harmonic provides a 'frame' for the attosecond 'movie' that lasts about half a laser cycle (here, about 1.4 fsec.)

High harmonic generation begins with strong-field ionization and ends with recombination. Orbitals relevant for ionization and recombination are the Dyson orbitals (see, for example, ref. 25): the overlaps between the multi-electron wavefunctions of the neutral and the ion. Figure 2a shows the Dyson orbitals corresponding to creating the ion CO_2^+ in the ground state $\tilde{X}\,^2\Pi_g$ (channel X), first excited state $\tilde{A}\,^2\Pi_u$ (channel A), and the second excited state $\tilde{B}\,^2\Sigma^+_u$

(channel B). In the molecular orbital picture, these channels describe the removal of an electron from HOMO, HOMO − 1 and HOMO − 2, correspondingly.

Consider the Dyson orbital for channel X (Fig. 2a). Ionization along or perpendicular to the molecular axis will be suppressed[19,20] owing to the destructive interference of currents from the orbital 'lobes' with opposite signs. This expectation is confirmed by the experiment[26] and our calculations (see 'Supplementary Information'). For channel A, the Dyson orbital favours ionization perpendicular to the molecular axis. For channel B, it favours ionization parallel to the molecular axis. Thus, rotating the molecule relative to the laser polarization controls the relative contributions of different orbitals. The importance of excited states is not specific to CO_2: they will be involved whenever the HOMO has nodal planes and/or when electronic excitations in the ion are close in energy.

Now, consider recombination. The contribution of each channel j with ionization potential $I_{p,j}$ to the Nth harmonic with energy $N\hbar\omega$ is proportional to the recombination matrix elements d at the electron energy $E_{e,j} = N\hbar\omega - I_{p,j}$. These are shown in Fig. 2b–d (see Supplementary Information for details). As a function of the alignment angle θ, both amplitudes and phases of d vary in opposite directions for HOMO versus HOMO − 2. The overall minimum in $|d|^2$ for channel X (HOMO) at small θ is not accidental: the same nodal plane in the Dyson orbital that suppresses ionization also suppresses recombination. These suppressions in channel X help channel B to make a comparable contribution to harmonic emission for small angles θ. Thus, measurement of the harmonic phases and amplitudes as a function of θ provides a diagnostic for the contribution of different orbitals.

The cutoff in the harmonic spectrum is determined by the maximum energy of the recombining electron plus the ionization potential I_p of the state into which it recombines[27]. Thus, contributions
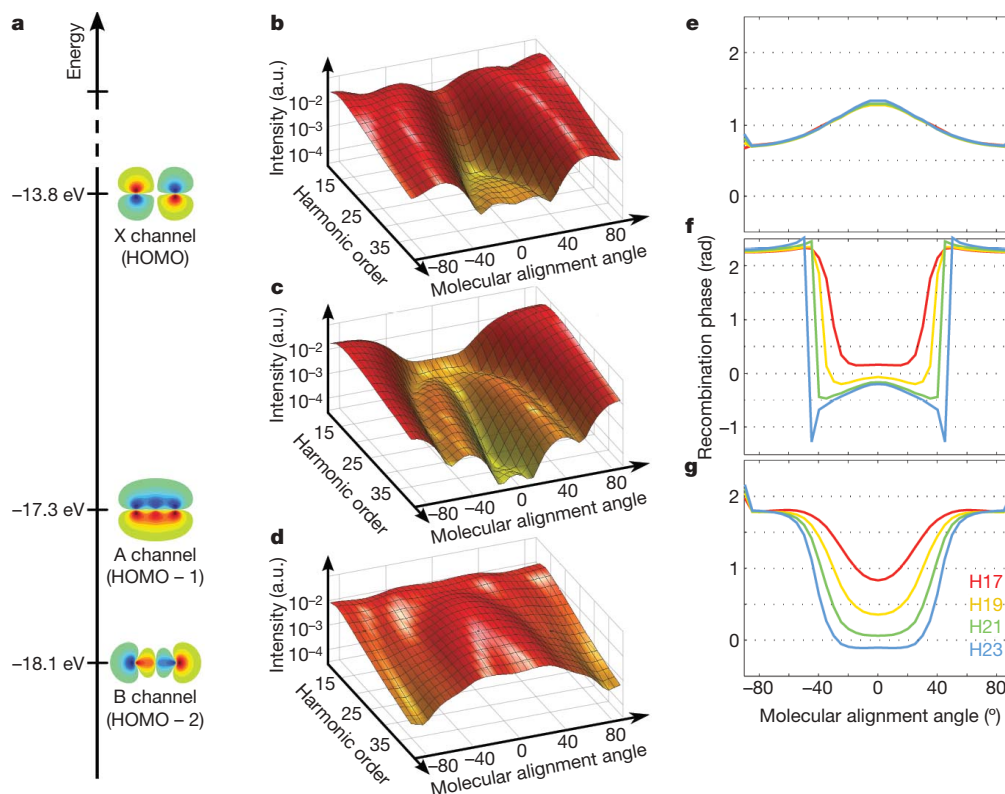
Figure 2 | Contributions of different molecular orbitals to the harmonic emission. **a**, Dyson orbitals corresponding to different final states of the CO_2^+ ion and the corresponding ionization energies. Differently coloured (blue and orange) orbital 'lobes' have opposite signs. **b–g**, Recombination amplitude squared (**b–d**) and recombination phase (**e–g**) show the calculated recombination matrix elements versus harmonic order and molecular alignment angle. The recombination phase (radians) is shown for selected harmonics H17, H19, H21 and H23. a.u., arbitrary units.

from deeper orbitals with higher I_p can be studied by looking beyond the cutoff of the harmonics associated with the HOMO: the B channel in CO_2 dominates the cutoff emission for molecules aligned at small angles θ, while the A channel is significant for perpendicular alignment, $\theta \approx 90°$.

Setup for high-harmonic interferometry

The main method of probing inner orbitals is to perform full characterization of the harmonic radiation, measuring both spectral phase and amplitude. To measure the dependence of the harmonic phase on the molecular alignment angle θ we use an approach pioneered by ref. 21. We create two spatially separated harmonic sources in a gas jet (Fig. 3a). One is a reference with randomly aligned molecules, while in the other molecular alignment is controlled. In the far field, we observe an interference pattern, similar to the one obtained with Young's slits[9,21,28,29]. To observe a stable interference pattern, the fluctuations of delay between the two generating laser beams must be smaller than the harmonic period (for example, 130 attoseconds for the 21st harmonic of a 800-nm laser pulse). We achieve a 1-attosecond stability by using a collinear scheme in which the two arms of the interferometer propagate together.

The position of the interference fringes is determined by the relative phase between the two sources. As the molecule is rotated away from the laser polarization, we observe a continuous shift of the fringes (Fig. 3b). Similar measurements for the N_2 molecules show the shift in the opposite direction, proving the ability of our setup to measure both the magnitude and the sign of the harmonic phase.

Identifying participating channels

The measured harmonic amplitude and phase versus alignment angle θ are shown in Fig. 3c, d. The phase variation is 0.4 rad for H17 and H19 and increases to 1.5 rad for H21 and 2 rad for H23. The curvature of the measured phase for H23 (Fig. 3d) is opposite to that calculated for the recombination matrix element in the X channel (HOMO) but similar to that for the B channel (Fig. 2e, g). This observation suggests that the harmonic generation in CO_2 cannot

be described assuming that only HOMO contributes to the process. The presence of the B channel is natural: at the laser intensity $I \approx 0.8 \times 10^{14}\ \mathrm{W\ cm^{-2}}$ used in these measurements H23 is already beyond the cutoff for the X channel.

Turning to the measured amplitudes (Fig. 3c), for harmonics 17 and 19 the signal is minimum at parallel alignment and maximum around 60°. Harmonics 23 and 25 (latter not shown) are maximal at parallel alignment. The former behaviour is consistent with the recombination matrix elements for the X channel, while the latter is consistent with the B channel (see Fig. 2). We note that for the X channel, both ionization and recombination are strongly suppressed at parallel alignment, while the observed minimum in the harmonic emission is rather modest. Therefore, experiment and theory indicate that the B channel (HOMO-2 orbital) is also contributing to the harmonics below H23 at this laser intensity. In this region the harmonic emission results from the interference of the contributions from different channels, as described in Fig. 1.

Channel interference and hole dynamics

We now turn to the analysis of the channel interference and the multi-electron dynamics it encodes. We show that this dynamics can be decoded by studying the interference pattern in the harmonic spectrum as a function of the laser intensity.

Experimentally, we have recorded the harmonic spectra in CO_2 molecules as a function of molecular alignment angle, for laser intensities from $I \approx 5 \times 10^{13}\ \mathrm{W\ cm^{-2}}$ to $I \approx 2 \times 10^{14}\ \mathrm{W\ cm^{-2}}$ (Fig. 4a, b and Fig. 5b, c). There is a clear minimum in the harmonic spectrum at $\theta = 0°$ (Fig. 4a, b), which shifts to higher orders with increasing intensity, from H19 at $0.55 \times 10^{14}\ \mathrm{W\ cm^{-2}}$ to H29 at $1.8 \times 10^{14}\ \mathrm{W\ cm^{-2}}$ (Fig. 5c). This observation explains why the minimum has been observed at vastly different positions in previous experiments[4,5,8,9]. The minimum has previously been associated with the structure of HOMO, but its mobility with intensity indicates that this picture is incomplete. Indeed, up to the details of the molecular alignment and saturation of ionization[17], the structural features are tied to the electron de Broglie wavelength and hence to a specific
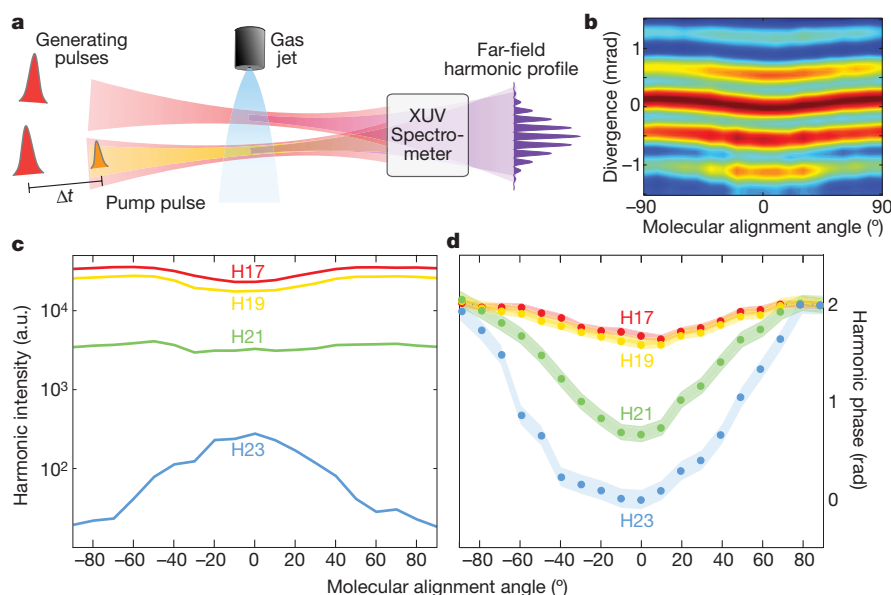
Figure 3 | Harmonic phase interferometry. a, Experimental setup. Two harmonic sources are created by focusing two synchronized non-collinear laser beams (probe) in a gas jet. A third laser beam (pump) is superimposed on the lower source to control molecular alignment. The pump–probe delay Δt is adjusted so that molecules are aligned along the pump pulse polarization when the generating pulses hit the jet. The harmonic radiation is analysed by an extreme ultraviolet (XUV) spectrometer that images the interference pattern of each harmonic. **b**, Measured interference fringes as a

function of molecular angle for the 21st harmonic in CO_2. The fringe pattern for each angle is averaged over 5,000 laser shots and normalized. **c, d**, Measured harmonic intensities (**c**) and phases (**d**) at $I \approx 8 \times 10^{13}\ \mathrm{W\ cm^{-2}}$. The phases of different harmonics are vertically shifted to coincide at 90°. The thickness of the line is the error bar, given by the standard deviation of the phase fluctuations in the absence of the pump beam (0.04 rad for H17).
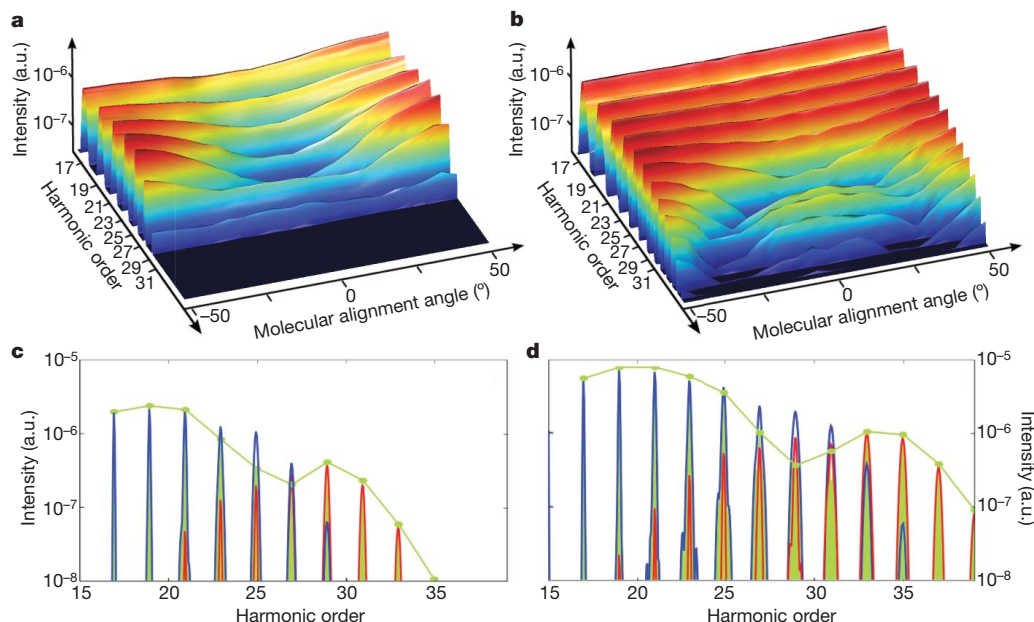
Figure 4 | Harmonic spectra in CO_2 molecules. a, b, Measured harmonic spectra as a function of the harmonic number and the molecular angle in CO_2, for $I \approx 1.1 \times 10^{14}\,W\,cm^{-2}$ (**a**) and $I \approx 1.8 \times 10^{14}\,W\,cm^{-2}$ (**b**). **c, d,** Calculated spectra for $\theta = 0°$ (after averaging over alignment distribution in the molecular ensemble), for $I = 1.26 \times 10^{14}\,W\,cm^{-2}$ (**c**) and

$I = 1.72 \times 10^{14}\,W\,cm^{-2}$ (**d**). The blue and red lines are the contributions from the channel X (HOMO) and channel B (HOMO − 2). The green line and the shaded spectrum is the total harmonic spectrum resulting from the coherent superposition of these contributions.
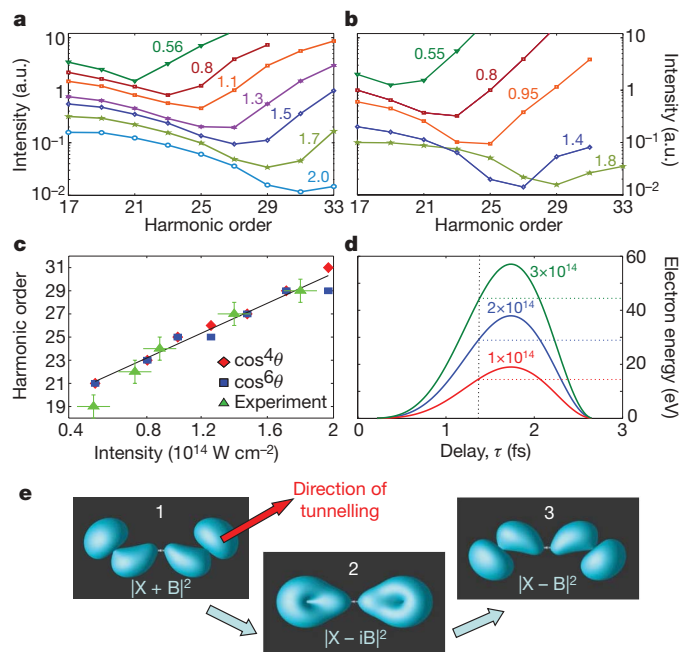
Figure 5 | Reconstruction of the hole dynamics. Calculated (**a**) and measured (**b**) spectra (integrated over each harmonic line) for different intensities and $\theta = 0°$ (shifted vertically for better visibility). Intensities are given in units of $10^{14}\,W\,cm^{-2}$. The spectra are normalized by the signal at $\theta = 50°$. **c,** Measured (triangles) and calculated intensity dependence of the position of the minimum. Calculations are for $\cos^4\theta$ (diamonds) and $\cos^6\theta$ (squares) alignment distributions. Error bars in the position of the minimum are based on standard deviation in the fluctuations of harmonic intensity. The line is a least-squares fit to the calculations. **d,** Origin of the intensity dependence for a dynamical minimum. For fixed emission time (horizontal axis) the electron energy E_e (vertical axis) and the harmonic order $N\hbar\omega = E_e + I_p$ linearly depend on the intensity (red, $I = 10^{14}\,W\,cm^{-2}$; blue, $I = 2 \times 10^{14}\,W\,cm^{-2}$; green, $I = 3 \times 10^{14}\,W\,cm^{-2}$). **e,** Visualization of the hole dynamics: (1) initial shape of the hole, (2) hole density quarter-period and (3) a half-period later.

harmonic order. The results of the previous section suggest that we are observing the destructive interference of channels X (HOMO) and B (HOMO − 2).

Our conclusion is supported by Fig. 4c and d, which shows the calculated spectra at $\theta = 0°$. The calculations (see Methods and Supplementary Information sections) include the contribution of X, A and B channels. To simulate experimental conditions, the contribution of the 'long trajectories' has been filtered out. Emitted radiation was coherently averaged over alignment distributions with parameters typical for the experiments. The X channel (blue curve) dominates the emission for lower harmonics, while the B channel (red curve) takes over for the higher harmonics. The total spectrum (green curve) records the relative phase between the channels by mapping it into the amplitude modulations. The minimum corresponds to the destructive interference between the channels.

The position of the minimum versus the laser intensity is shown in Fig. 5. Both theoretical (Fig. 5a) and experimental (Fig. 5b) spectra have been normalized to those for $\theta = 50°$. At this angle, channel X dominates the spectrum, channel interference is negligible, and the spectrum has no pronounced minima or maxima. Normalization eliminates the instrument response effects and makes the quantitative comparison of theory and experiment possible. Experiment and theory are in good agreement.

Why does the minimum shift with intensity and what can we learn from it? There are four different sources of relative phase between the channels, which control their interference (see equation (1)): (1) the recombination phase (Fig. 2e–g); (2) the continuum evolution phase, which is dominated by the strong laser field and is similar for different channels (see 'Supplementary Information'); (3) the evolution of the electronic states j of the ion $\Psi_j(t) \propto e^{-iE_j\tau}$ during the time-interval τ between ionization and recombination (E_j are the state energies); and (4) the ionization phase (see below).

The interference minimum is tied to that time-delay τ^* when channels interfere destructively. For a given τ^*, the energy of the continuum electron $E_e(\tau^*)$ varies linearly with the laser intensity (Fig. 5d). Therefore, in first approximation, the minimum should move linearly with intensity. A weaker influence of the laser field on the bound

975

electrons in the $CO_2{}^+$ ion is also included in our calculations (see Supplementary Information).

Experiment confirms that the harmonic minimum shifts approximately linearly with intensity. A least-squares fit to the data (Fig. 5c) shows that the position of the minimum scales as $N\hbar\omega \propto (1.7 \pm 0.1)U_p$, where U_P is the mean quiver energy of the electron in the laser field. The mapping of the harmonic numbers onto the time-delay τ between ionization and recombination[6,7,23,24,30], first used to track the motion of protons[6,7,24], yields $\tau^* = 1.17 \pm 0.1$ fsec after ionization. At this moment the relative phase accumulated in the ion is $(E_B - E_X)\tau^* \approx (2.44 \pm 0.2)\pi$. Destructive interference requires a total phase difference of 3π. The phase between the channels also includes the relative phase between the recombination matrix elements, which is about 0.5π (note the opposite parity of HOMO and HOMO $- 2$). Thus, the total relative phase between the channels is indeed 3π at τ^*, demonstrating the consistency of our reconstruction. Additionally, reconstruction of τ^* from the position of the minimum for each intensity separately yields the same result.

Coherent population of two ionic states (for example, X and B) implies excitation of the bound electronic wavepacket in the ion. Interference between the channels records the wavepacket dynamics. This two-state dynamics has one non-trivial parameter—the initial relative phase φ between the two ionic states that could have been acquired during ionization. In the static tunnelling limit, in each channel the wavefunction after electron tunnelling through the barrier is an extension of the real wavefunction that characterizes the bound state, implying $\varphi = 0$. Non-zero φ would reflect non-adiabatic dynamics of electron rearrangement during strong-field ionization. We note that φ is different from the phase of the dipole transition associated with recombination. Indeed, recombination is a single-photon transition, whereas optical tunnelling proceeds via absorption of many infrared photons.

Knowing the instant of destructive interference allows us to find φ. At high intensities, the reconstructed value $\tau^* = 1.17 \pm 0.1$ fsec is consistent with $\varphi = 0$, expected in the tunnelling regime. Changing φ to $\varphi = \pm\pi/2$ in the theoretical calculations moves the calculated positions of the minimum beyond the experimental error; see Supplementary Information.

For the same state of the continuum electron, the initial phase between the channels determines the initial location/shape of the hole left in the molecule by ionization. It could be visualized as a superposition of the Dyson orbitals $\Psi^{(D)}{}_X$ and $\Psi^{(D)}{}_B$ corresponding to channels X and B: $|\Psi_{init}|^2 = |\Psi^{(D)}{}_X + e^{+i\varphi}\Psi^{(D)}{}_B|^2$, for $\varphi = 0$ (see Fig. 5e, snapshot 1). The phase $\varphi = 0$ yields the maximum possible extension of the density in the direction of tunnelling (for the two channels X and B.) Snapshots 2 and 3 in Fig. 5e show the hole density quarter-period later $(|\Psi^{(D)}{}_X - i\Psi^{(D)}{}_B|^2)$ and half-period later $(|\Psi^{(D)}{}_X - \Psi^{(D)}{}_B|^2)$.

Perspectives

Our work shows how measurements of the phases and amplitudes of high harmonic radiation can be used to follow not only vibrational[6,7], but also multi-electron dynamics in molecules with attosecond temporal resolution. These dynamics can be induced by ionization, as in the CO_2 molecule. Multi-electron excitations can also be induced by the laser field between ionization and recombination, on the sub-cycle timescale[31–33]. Multiple ionic states are naturally excited during photo-ionization and/or non-adiabatic transitions induced by nuclear motion. High harmonic emission, used as a probe in ultrafast pump–probe experiments (see, for example, refs 10 and 34), provides a new tool with which to measure the underlying dynamics.

The question of the relative phase φ between different channels in strong-field ionization has not been addressed before. This phase records interactions between the departing electron and the electrons of the core, that is, the coupling between different channels during strong-field ionization. We expect that the phase φ will depend on the ionization regime, because the coupling between the channels during

strong-field ionization is caused not noly by electron–electron interaction, but also laser field. Analysis of this dependence based purely on the experimental measurements would require a separate characterization of the phase of recombination, which should be possible using interferometric techniques of attosecond photoelectron spectroscopy (see for example, refs 35 and 36). Our experiment (Fig. 5c) seems to suggest that in the multiphoton regime (below $I \approx 1.1 - 1.2 \times 10^{14}$ W cm^{-2}), the intensity dependence of the minimum might be deviating from the linear; see Supplementary Information for discussion. However, the deviation is at the limits of the accuracy of our measurement. Thus, high-harmonic interferometry offers a versatile approach to observing a wide variety of attosecond processes. Combined with the recently developed perturbative methods for attosecond metrology[30] and the measurements of the polarization state[37], it will greatly facilitate a new frontier of attosecond dynamics measurements.

METHODS SUMMARY

Experimentally, we use a 50-Hz Ti:Sa laser system providing 15-mJ 30-fs pulses at 800 nm. The laser pulse is split by a polarizing beam splitter and a half-waveplate. The transmitted beam is the alignment beam. Its polarization direction is rotated with a second half-waveplate. Critical to the experiment is a very stable interferometer that is formed by splitting the reflected beam into two generating beams. We exploit the refractive index difference between two axes of a birefringent crystal to produce the two beams, separated by a small angle. Both beams share all optical elements. The three beams are focused by a 50-cm lens in a 1-mm-long pulsed gas jet. The high harmonics produced by the two generating beams, separated by $\sim 70\,\mu m$, overlap in the far field, which serves as a beam combiner. The harmonic intensity is obtained by integrating the far-field spatial profile, while the harmonic phase is extracted by means of the Fourier transform.

Theoretically, multichannel harmonic emission is described by the induced dipole:

$$\mathbf{D}(t) \approx \langle \Psi_{NT}(t)|\hat{\mathbf{d}}| \sum_j \hat{\mathbf{A}}\Psi_j(t)\chi_{C,j}(t)\rangle + c.c. \qquad (1)$$

Here $\Psi_{NT}(t)$ describes the evolution of the neutral molecule during the laser pulse, including depletion by ionization, while the electron-ion wavepacket created by ionization is $\Psi_{ei}(t) = \sum_j \hat{\mathbf{A}}\Psi_j(t)\chi_{C,j}(t)$. Index j labels the channels involving different ionic states, $\Psi_j(t)$ describes their evolution between ionization and recombination, $\chi_{C,j}(t)$ are the continuum wavepackets correlated to these ionic states, the operator $\hat{\mathbf{A}}$ anti-symmetrizes the electrons, and $c.c.$ indicates the complex conjugate. The field of Nth harmonic is determined by the Fourier transform $\tilde{\mathbf{D}}(N\omega)$ of $\mathbf{D}(t)$ at the frequency $N\omega$. The Fourier transform enforces the energy conservation law, $N\hbar\omega = E_e + I_{p,j}$, where E_e is the energy of the returning electron and $I_{p,j}$ is the ionization potential in channel j.

1. Cohen, H. D. & Fano, U. Interference in the photo-ionization of molecules. *Phys. Rev.* **150**, 30–33 (1966).
2. Lein, M. *et al.* Role of the intermolecular phase in high harmonic generation. *Phys. Rev. Lett.* **88**, 183903 (2002).
3. Itatani, J. *et al.* Tomographic imaging of molecular orbitals. *Nature* **432**, 867–871 (2004).
4. Kanai, T., Minemoto, S. & Sakai, H. Quantum interference during high-order harmonic generation from aligned molecules. *Nature* **435**, 470–474 (2005).
5. Vozzi, C. *et al.* Controlling two-center interference in molecular high harmonic generation. *Phys. Rev. Lett.* **95**, 153902 (2005).
6. Baker, S. *et al.* Probing proton dynamics in molecules on an attosecond time scale. *Science* **312**, 424–427 (2006).
7. Baker, S. *et al.* Dynamic two-center interference in high-order harmonic generation from molecules with attosecond nuclear motion. *Phys. Rev. Lett.* **101**, 053901 (2008).
8. Boutu, W. *et al.* Coherent control of attosecond emission from aligned molecules. *Nature Phys.* **4**, 545–549 (2008).
9. Zhou, X. *et al.* Molecular recollision interferometry in high harmonic generation. *Phys. Rev. Lett.* **100**, 073902 (2008).
10. Li, W. *et al.* Time-resolved dynamics in N_2O_4 probed using high harmonic generation. *Science* **322**, 1207–1211 (2006).
11. McFarland, B. K. *et al.* High harmonic generation from multiple orbitals in N_2. *Science* **322**, 1232–1235 (2006).
12. Rosca-Pruna, F. & Vrakking, M. J. J. Experimental observation of revival structures in picosecond laser-induced alignment of I_2. *Phys. Rev. Lett.* **87**, 153902 (2001).

13. Stapelfeldt, H. & Seideman, T. Aligning molecules with strong laser pulses. *Rev. Mod. Phys.* **75**, 543–557 (2003).

14. Le, A. T., Morishita, T. & Lin, C. D. Extraction of the species-dependent dipole amplitude and phase from high-order harmonic spectra in rare-gas atoms. *Phys. Rev. A* **78**, 023814 (2008).

15. Morishita, T., Le, A. T., Chen, Z. & Lin, C. D. Accurate retrieval of structural information from laser-induced photoelectron and high-order harmonic spectra by few-cycle laser pulses. *Phys. Rev. Lett.* **100**, 013903 (2008).

16. Le, V. H., Le, A. T., Xie, R. H. & Lin, C. D. Theoretical analysis of dynamic chemical imaging with lasers using high-order harmonic generation. *Phys. Rev. A* **76**, 013414 (2007).

17. Le, A. T., Tong, X. M. & Lin, C. D. Evidence of two-center interference in high-order harmonic generation from CO_2. *Phys. Rev. A* **73**, 041402 (2006).

18. Lein, M. Molecular imaging using recolliding electrons. *J. Phys. B* **40**, R135–R173 (2007).

19. Tong, X. M., Zhao, Z. X. & Lin, C. D. Theory of molecular tunneling ionization. *Phys. Rev. Lett.* **66**, 033402 (2002).

20. Muth-Böhm, J., Becker, A. & Faisal, F. H. M. Suppressed molecular ionization for a class of diatomics in intense femtosecond laser fields. *Phys. Rev. Lett.* **85**, 2280–2283 (2000).

21. Zerne, R. et al. Phase locked high-order harmonic emission. *Phys. Rev. Lett.* **79**, 1006–1009 (1997).

22. Corkum, P. B. Plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).

23. Mairesse, Y. et al. Attosecond synchronization of high-harmonic soft X-rays. *Science* **302**, 1540–1543 (2003).

24. Lein, M. Attosecond probing of vibrational dynamics with high-harmonic generation. *Phys. Rev. Lett.* **94**, 053004 (2005).

25. Ortiz, J. V. Quasiparticle approximations and electron propagator theory. *Int. J. Quantum Chem.* **95**, 593–599 (2003).

26. Pavicic, D. et al. Direct measurement of the angular dependence of ionization for N_2, O_2, and CO_2 in intense laser fields. *Phys. Rev. Lett.* **98**, 243001 (2007).

27. Lewenstein, M. et al. Theory of high harmonic generation by low-frequency laser fields. *Phys. Rev. A* **49**, 2117–2132 (1994).

28. Descamps, D. et al. Extreme ultraviolet interferometry measurements with high order harmonics. *Opt. Lett.* **25**, 135–137 (2000).

29. Corsi, C., Pirri, A., Sali, E., Tortora, A. & Bellini, M. Direct interferometric measurement of the atomic dipole phase in high-order harmonic generation. *Phys. Rev. Lett.* **97**, 023901 (2006).

30. Dudovich, N. et al. Measuring and controlling the birth of attosecond XUV pulses. *Nature Phys.* **2**, 781–786 (2006).

31. Lezius, M. et al. Nonadiabatic multielectron dynamics in strong field molecular ionization. *Phys. Rev. Lett.* **86**, 51–89 (2001).

32. Markevitch, A. N. et al. Nonadiabatic dynamics of polyatomic molecules and ions in strong laser fields. *Phys. Rev. A* **68**, 011402 (2003).

33. Litvinyuk, I. et al. Shakeup excitation during optical tunnel ionization. *Phys. Rev. Lett.* **94**, 033003 (2005).

34. Wagner, N. et al. Monitoring molecular dynamics using coherent electrons from high-harmonic generation. *Proc. Natl Acad. Sci. USA* **103**, 13279–13285 (2006).

35. Remetter, T. et al. Attosecond electron wave packet interferometry. *Nature Phys.* **2**, 323–326 (2006).

36. Corkum, P. B. & Krausz, F. Attosecond science. *Nature Phys.* **3**, 381–387 (2007).

37. Levesque, J. et al. Polarization state of high-order harmonic emission from aligned molecules. *Phys. Rev. Lett.* **99**, 243001 (2007).

Author Contributions O.S., S.P. and M.Y.I. developed theory, and Y.M., N.D., D.V. and P.C. performed experiments. All authors contributed to the analysis of data and writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.S. (olga.smirnova@mbi-berlin.de).

nature

ARTICLES

Homotypic fusion of ER membranes requires the dynamin-like GTPase Atlastin

Genny Orso[1]*, Diana Pendin[1]*, Song Liu[2]*, Jessica Tosetto[1,3], Tyler J. Moss[2], Joseph E. Faust[2], Massimo Micaroni[4], Anastasia Egorova[5], Andrea Martinuzzi[1], James A. McNew[2] & Andrea Daga[1,6,7]

Establishment and maintenance of proper architecture is essential for endoplasmic reticulum (ER) function. Homotypic membrane fusion is required for ER biogenesis and maintenance, and has been shown to depend on GTP hydrolysis. Here we demonstrate that *Drosophila* Atlastin—the fly homologue of the mammalian GTPase atlastin 1 involved in hereditary spastic paraplegia—localizes on ER membranes and that its loss causes ER fragmentation. *Drosophila* Atlastin embedded in distinct membranes has the ability to form *trans*-oligomeric complexes and its overexpression induces enlargement of ER profiles, consistent with excessive fusion of ER membranes. *In vitro* experiments confirm that Atlastin autonomously drives membrane fusion in a GTP-dependent fashion. In contrast, GTPase-deficient Atlastin is inactive, unable to form *trans*-oligomeric complexes owing to failure to self-associate, and incapable of promoting fusion *in vitro*. These results demonstrate that Atlastin mediates membrane tethering and fusion and strongly suggest that it is the GTPase activity that is required for ER homotypic fusion.

The ER is composed of a network of interconnected tubules that pervades the cytoplasm of eukaryotic cells. Homotypic membrane fusion is essential for ER establishment and maintenance[1–3]. This activity requires GTP hydrolysis and non-cytosolic factors[1,2,4], suggesting the involvement of an as yet unidentified GTP-dependent fusion machinery associated with the ER membrane. Human atlastin GTPase 1 (ATL1), the mutation of which causes a form of hereditary spastic paraplegia, belongs to the dynamin superfamily of GTPases[5] and has been implicated in ER–Golgi vesicle trafficking and Golgi morphogenesis[6,7]. Two other atlastins[7] have been identified in mammals, but their function remains unexplained. *Drosophila* contains a single highly conserved atlastin orthologue (Atlastin). The *atlastin*-null mutant flies (*atl[1]*) are abnormally small and have motor dysfunction[8], however the cellular role of Atlastin has not been explored. Here we report that Atlastin localizes to ER membranes and that the loss of Atlastin function causes fragmentation of the ER network. Atlastin localized on adjacent membranes can form *trans*-oligomeric complexes, and its overexpression results in the formation of expanded ER elements, consistent with excessive fusion of membranes. Notably, Atlastin is autonomously capable of promoting *in vitro* fusion of liposomes. In contrast, the expression of GTPase-deficient Atlastin has no phenotypic consequences, cannot form *trans*-oligomeric complexes, and fails to drive membrane fusion *in vitro*. These results indicate that Atlastin is the crucial GTPase mediating homotypic fusion of ER membranes.

Atlastin localizes to the ER

Drosophila and human Atlastins show remarkable homology and conservation of domain organization (Supplementary Fig. 1), probably resulting in analogous membrane topology[7,9]. Immunohistochemistry experiments showed that *Drosophila* Atlastin is ubiquitously expressed, and its expression levels are high during embryonic development. In

embryos, Atlastin immunoreactivity consistently overlapped with the ER-reporter green fluorescent protein (GFP)–KDEL and the ER-localized proteins BiP and Rtnl1–GFP[10] (Fig. 1a, b). Simultaneous visualization of endogenous Atlastin and the Golgi apparatus using the Golgi reporter GalT–GFP[11] or antibodies to the Golgi marker p120 (ref. 12) demonstrate that Atlastin does not share the punctate distribution that is typical of Golgi in *Drosophila*[13] (Fig. 1a, b). Moreover, quantitative immunogold electron microscopy (EM) analysis in *Drosophila* S2 cells showed that approximately 60% of gold particles were found on ER membranes (Supplementary Fig. 2). Together, these results establish that Atlastin is a membrane protein residing in the ER.

Loss of Atlastin causes ER fragmentation

Secretory pathway traffic has been reported to be normal under knockdown conditions for all human atlastins[9]. Similarly, we found that the loss of *Drosophila* Atlastin did not alter mCD8–GFP membrane transport, had no effects on Golgi morphology and distribution, and did not perturb ER exit sites, indicating that Atlastin is unlikely to participate in secretory traffic (Supplementary Information and Supplementary Figs 3–5). To gain insight into the function of Atlastin, we examined the consequences of downregulating its expression in *Drosophila* by *in vivo* RNA interference (RNAi). We generated UAS-*atl*-RNAi transgenic flies, the expression of which can be controlled using the Gal4-UAS system[14]. Several experiments demonstrated that *in vivo atlastin* knockdown was specific and effective (Supplementary Information and Supplementary Fig. 6). Loss of Atlastin in muscle and neurons resulted in modest changes in GFP–KDEL fluorescence (Supplementary Fig. 7, Supplementary Movies 1, 2); however, ultrastructural analysis of ER morphology showed considerable alterations (Fig. 2). Control neurons had long tubular ER profiles (average length 876 nm), whereas neurons lacking

---

[1]Eugenio Medea Scientific Institute, Conegliano 31015, Italy. [2]Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005, USA. [3]Department of Pharmacology, University of Padova, Padova 35131, Italy. [4]University of Queensland, Institute for Molecular Bioscience, St Lucia, Brisbane, Queensland 4072, Australia. [5]Department of Cell Biology and Oncology, Consorzio "Mario Negri Sud", Santa Maria Imbaro 66030, Italy. [6]Dulbecco Telethon Institute, Eugenio Medea Scientific Institute, Padova 35131, Italy. [7]Department of Neurology, The David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.
*These authors contributed equally to this work.
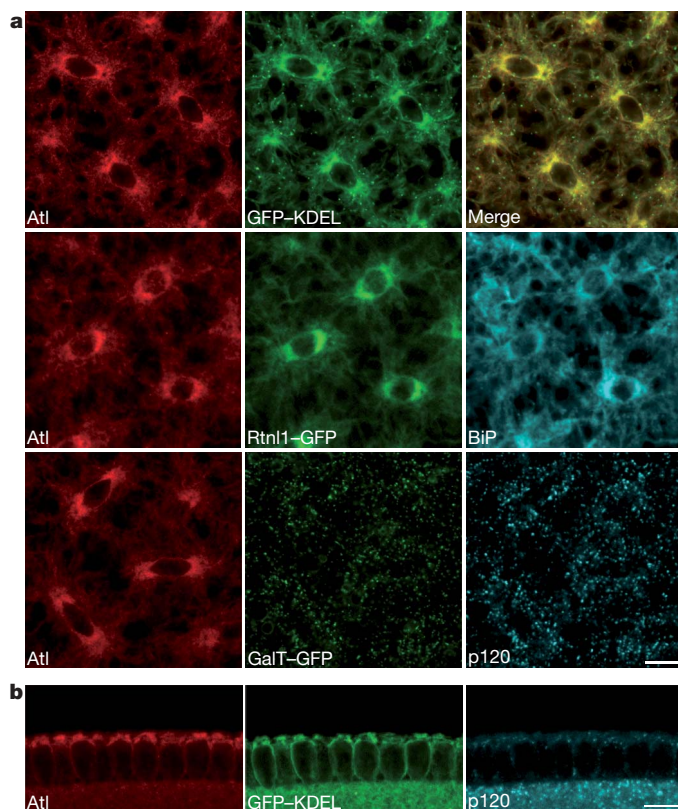
Figure 1 | Endogenous Atlastin localizes on the ER. a, In syncytial blastoderm embryos, anti-Atlastin immunofluorescence (red) overlaps with that of the ER markers GFP–KDEL (green) (top row), Rtnl1–GFP (green) and BiP (cyan) (middle row). Visualization of Golgi apparatus with the GalT–GFP reporter (green) and anti-p120 antibody (cyan) shows that Atlastin immunoreactivity (red) does not co-localize with Golgi punctae (bottom row). Scale bar, 10 µm. **b,** Sagittal view of a cellularizing embryo labelled with anti-Atlastin antibody (red), GFP–KDEL (green) and anti-p120 antibody (cyan). Atlastin immunoreactivity overlaps with GFP–KDEL fluorescence but not with the p120 signal. Scale bar, 10 µm.

Atlastin showed severely undersized ER profiles (average length 308 nm) (Fig. 2a, b). Furthermore, the ER profile size distribution in *atlastin* RNAi neurons showed two classes of short ER profiles (0–200 and 200–400 nm) that were virtually absent from the cytoplasm of wild-type neurons (Fig. 2c). EM analysis showed analogous shortened ER profiles in *atl*[1] mutant neurons[8] (Fig. 2a–c), and in tubulin-Gal4/+;UAS-*atl*-RNAi/+ muscles (data not shown). In addition, the expression of Atlastin in the null mutant background fully rescued ER fragmentation (Supplementary Fig. 8). These results indicate that the ER is fragmented in response to loss of Atlastin.

The functional consequences of these morphological changes in ER structure were examined by fluorescence loss in photobleaching (FLIP). The loss of a fluorescent marker from a region of a cell after repeated photobleaching of a different region indicates continuity between the two regions. We used FLIP experiments targeting GFP–KDEL[11,15] to assess whether fragmentation in *atlastin* RNAi tissue resulted in discontinuity of normally interconnected ER elements. Unlike in control muscle where the loss of GFP–KDEL fluorescence was homogeneous, repetitive photobleaching of GFP–KDEL in tubulin-Gal4/+;UAS-*atl*-RNAi/+ muscle produced regions of unbleached fluorescence (Fig. 2d, e), indicating that in these areas the ER network lacked its typical continuity. Fluorescence loss was still detectable in other areas (Fig. 2d, e, region of interest (ROI) 3 in RNAi) suggesting that fragmentation was partial, probably because it is the result of deterioration of ER maintenance (Supplementary Information). These data reinforce our EM observations demonstrating that the removal of Atlastin results in ER fragmentation.

Atlastin overexpression results in expanded ER membranes

Because a reduction in Atlastin levels produced ER fragmentation, we predicted that the overexpression of Atlastin may lead to excessive membrane fusion if Atlastin were involved in this process. Immunofluorescence analysis of larva brains overexpressing Atlastin–Myc with the motoneuron driver D42-Gal4 (ref. 16) showed that the ER marker BiP accumulated in cytoplasmic structures that were absent in controls. These structures contained Atlastin and the Golgi marker p120 (Fig. 3a), suggesting that Atlastin–Myc localizes properly but alters ER morphology and disrupts the Golgi apparatus. EM analysis of D42-Gal4/+;UAS-*atlastin-myc*/+ brains showed that Atlastin overexpression disrupts the ER network. Normal tubular ER profiles were absent in Atlastin-overexpressing motoneurons where ER membranes formed expanded cisternae (Fig. 3b). This expansion of ER elements is consistent with an increase in membrane fusion and suggests that Atlastin itself could directly mediate bilayer merger. In agreement with immunofluorescence data (Fig. 3a), normal Golgi complexes were essentially absent in neurons overexpressing Atlastin. The absence of normal Golgi and the redistribution of Golgi proteins to the ER (Fig. 3a) suggest a block in secretory traffic[17], indicating that hyperfusion of ER membranes impairs this process (Supplementary Information and Supplementary Fig. 5).

Atlastin mediates tethering of ER membranes

Because human atlastin proteins self-assemble into oligomeric complexes[7,9], we sought to establish whether Atlastin was also capable of homo-oligomerization. HeLa cells were simultaneously transfected with Atlastin–Myc and Atlastin–haemagglutinin (HA) constructs, and co-immunoprecipitation experiments demonstrated self-association of Atlastin molecules (Fig. 3c and Supplementary Fig. 9). This finding, together with the observations that human atlastins are transmembrane proteins[7,9] and that Atlastin localizes to ER membranes and its downregulation triggers ER fragmentation, raised the possibility that Atlastin could be directly involved in tethering adjacent ER membranes, thereby permitting homotypic fusion to occur. To test this hypothesis, membrane vesicles were prepared from HeLa cells separately transfected with Atlastin–Myc or Atlastin–HA. Transfected cells were homogenized in the absence of detergent and fragmented membranes were vesiculated by sonication[7,18], mixed and immunoprecipitated. When anti-Myc antibodies were used to precipitate Atlastin–Myc-containing vesicles, Atlastin–HA was recovered in the pellet (Fig. 3d). Fractionation of cleared cell homogenates showed that Atlastin and the ER resident integral membrane protein calnexin partitioned exclusively to the membrane fraction, whereas the ER-luminal protein PDI remained in the soluble fraction (Fig. 3e). This demonstrates that under these lysis conditions, Atlastin associated with membranes and Atlastin binding occurred exclusively between distinct vesicles. Thus, Atlastin molecules inserted in adjacent ER membranes can form *trans*-complexes. These results establish that homophilic interactions between Atlastin molecules provide a tethering step between opposing ER membranes, leading to bilayer merger potentially mediated directly by Atlastin.

Atlastin drives membrane fusion *in vitro*

A direct role for Atlastin in membrane fusion was tested *in vitro*. Recombinant glutathione *S*-transferase (GST)–Atlastin was reconstituted into two populations of preformed liposomes[19] and fusion was measured by lipid-mixing[20,21]. Acceptor proteoliposomes contained unlabelled lipids and donor proteoliposomes contained these lipids as well as 7-nitrobenzoxadiazole (NBD) and rhodamine head-group-labelled lipids. NBD and rhodamine form a fluorescence resonance energy transfer (FRET) pair, and fusion is measured as an increase in NBD fluorescence over time as lipid mixing between donor and acceptor proteoliposomes dilutes the fluorescence probes in the newly merged membrane. When equimolar amounts of fluorescently labelled and unlabelled Atlastin proteoliposomes are mixed,
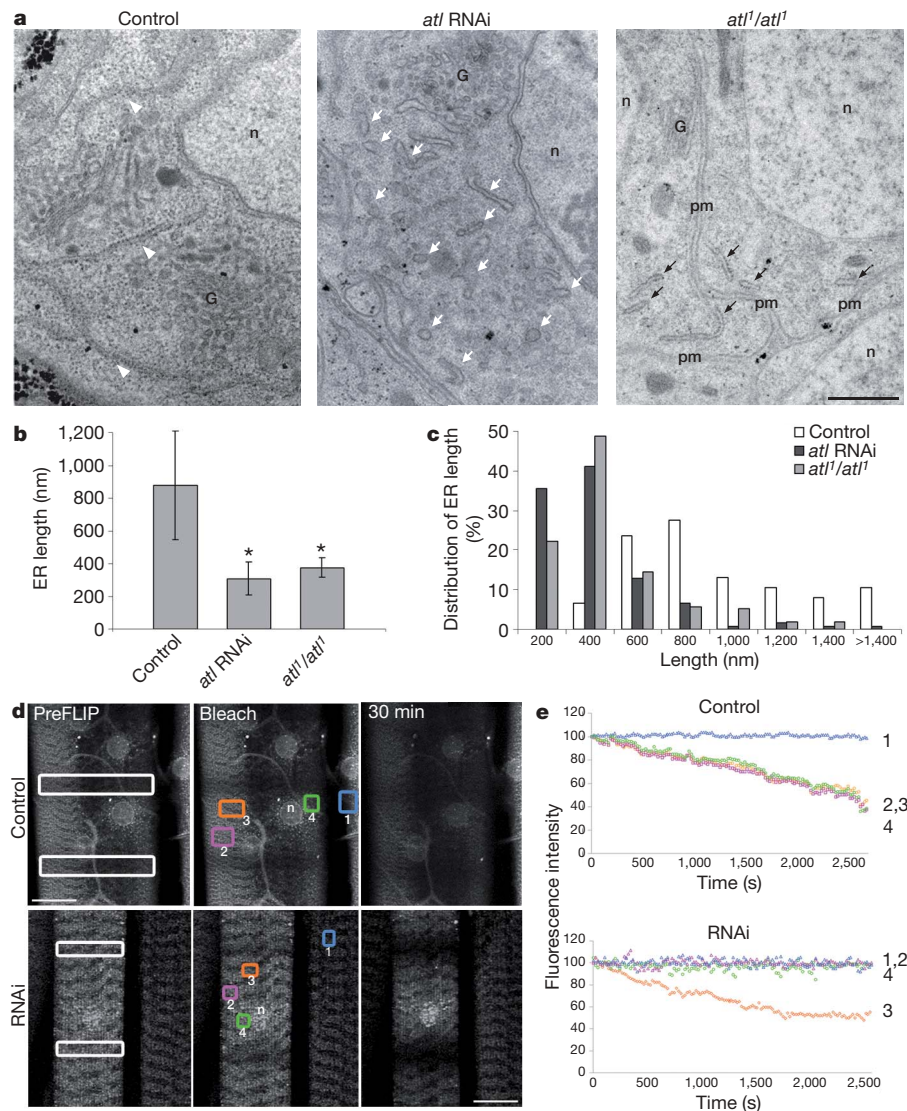
Figure 2 | Loss of Atlastin causes fragmentation and discontinuity of the ER. a, EM analysis shows that in control neurons the ER shows the typical tubular structure (arrowheads), whereas RNAi and *atl1* neurons exhibit a fragmented ER morphology (arrows). Scale bar, 0.5 μm. G, Golgi; n, nucleus; pm, plasma membrane. **b**, Average length of ER profiles. Error bars represent s.d.; $n > 100$; *$P < 0.000001$. **c**, Difference in size distribution of ER profiles. **d**, Repetitive photobleaching of two ROI (white outline boxes) in control and *atl* RNAi muscle expressing GFP–KDEL was performed. Fluorescence loss was analysed in four independent regions of the muscle (coloured outline boxes). ROI 1 was chosen on an adjacent unbleached muscle as a control. Scale bar, 20 μm. **e**, Rates of loss of fluorescence for individual ROI.

magnesium and GTP addition drives robust fusion that is completely dependent on GTPase activity (Fig. 4a and Supplementary Fig. 10) and independent of lipid composition (Supplementary Fig. 11). No fusion is seen using GDP, GMP, ATP or when magnesium is replaced with calcium (Supplementary Fig. 12), and fusion driven by Atlastin results in complete mixing of both inner and outer phospholipid monolayers (Supplementary Fig. 13). Atlastin-mediated fusion is concentration-dependent (Fig. 4b) and correlates with an increase in size of the proteoliposome population as measured by dynamic light scattering (Fig. 4c, d and Supplementary Fig. 14). Furthermore, analysis of negative stained Atlastin proteoliposomes by EM showed a homogenous population of liposomes before and after fusion (Supplementary Fig. 15). These data demonstrate that Atlastin alone is sufficient to drive membrane fusion *in vitro*.

Atlastin function requires GTPase activity

We postulated that Atlastin-mediated tethering and fusion of ER membranes should rely on the GTPase activity of Atlastin. To test this hypothesis, we generated transgenic flies for the expression of Atlastin carrying a Lys51Ala substitution (UAS-*atlastin-(K51A)–myc*) in the

P-loop of the GTP-binding domain. Biochemical analysis of recombinant wild-type and mutant (Lys51Ala) Atlastin indicates that replacement of Lys 51 significantly reduces GTPase activity[22] (Fig. 5i). Consistent with this notion, GMR-Gal4 driven eye-specific expression of UAS-*atlastin-(K51A)–myc* had no phenotypic effects, whereas the expression of UAS-*atlastin–myc* gave rise to a small eye (Fig. 5a–c). Moreover, in contrast to overexpression of wild-type Atlastin, overexpression of Atlastin(K51A) with all drivers allowed survival of the flies, indicating that replacement of Lys 51 with Ala results in the inactivation of the protein. Notably, immunofluorescence (Fig. 5d) and ultrastructural (data not shown) analyses of motor neurons expressing Atlastin(K51A)–Myc under the control of D42-Gal4 showed that the mutant properly localized to the ER, and ER and Golgi morphology was unaltered. Moreover, Atlastin(K51A) was unable to rescue ER fragmentation in *atl1* neurons (Supplementary Fig. 8). These observations demonstrate that formation of aberrant ER depends crucially on the GTPase activity of Atlastin.

To understand how the loss of GTPase activity might impair Atlastin function, we examined the ability of the GTPase-deficient Atlastin(K51A)–HA and Atlastin(K51A)– Myc to homo-oligomerize
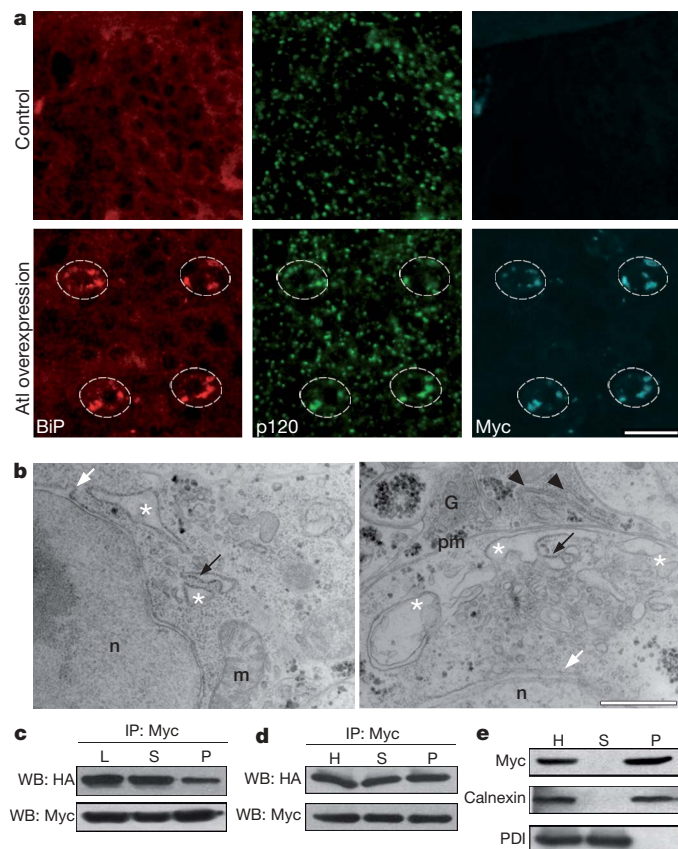
Figure 3 | Atlastin induces hyperfusion of ER membranes, homo-oligomerization and membrane tethering. a, Overexpression of Atlastin–Myc in motor neurons (white outline) induces the formation of cytoplasmic bodies with ER and Golgi identity, as shown by labelling with anti-Myc, anti-BiP and anti-p120 antibodies. Scale bar, 10 μm. **b,** EM analysis shows that neurons overexpressing Atlastin exhibit ER (asterisks) and nuclear envelope (white arrows) expansion. Note a non-overexpressing neuron with normal tubular ER morphology (black arrowheads). Scale bar, 0.5 μm. m, mitochondrion; n, nucleus; pm, plasma membrane. Black arrows denote ribosomes. **c,** Lysates from HeLa cells co-transfected with Atlastin–Myc and Atlastin–HA were immunoprecipitated (IP) and analysed by western blot (WB). **d,** Membrane vesicles from HeLa cells expressing Atlastin–HA or Atlastin–Myc were mixed, immunoprecipitated, and analysed by western blotting. **e,** Western blot analysis of the soluble and membrane fractions from cell homogenates. H, cell homogenate; L, lysate; P, pellet; S, supernatant.

in co-transfected HeLa cells as well as in the membrane-vesicle immunoprecipitation assay. In both assays, Atlastin(K51A) was unable to self-associate, indicating that this inability prevented the formation of *trans*-oligomeric complexes between Atlastin(K51A)–HA- and Atlastin(K51A)–Myc-containing vesicles (Fig. 5e–g). Thus, GTPase activity is critical for self-association and GTPase-deficient Atlastin(K51A) lacks the competence to mediate membrane tethering.

Given that Atlastin(K51A) was unable to oligomerize, we tested its ability to drive fusion. Donor and acceptor proteoliposomes containing equal amounts of Atlastin(K51A) or wild-type Atlastin (Supplementary Fig. 16) were prepared and their ability to fuse was analysed. When wild-type Atlastin was incorporated in both membranes, fusion proceeded normally; however, the inclusion of Atlastin(K51A) in the reaction failed to support fusion (Fig. 5h). This result demonstrates that a functional GTPase domain is required on both membranes for fusion to occur.

Discussion

Establishment of the ER network occurs by a basic homotypic fusion reaction that requires GTP hydrolysis and membrane-associated factors[1,2,4,23]. The formation of a tubular network then ensues that
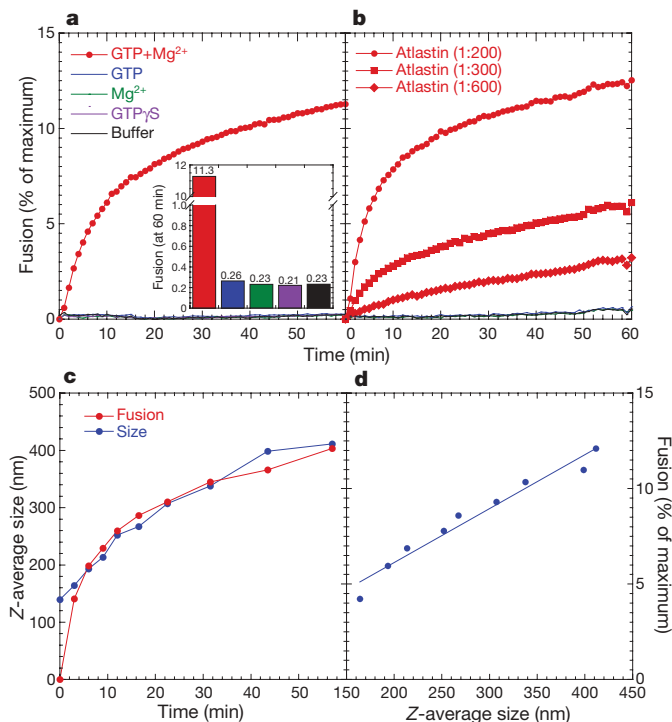
Figure 4 | Atlastin mediates proteoliposome fusion. a, Kinetic fusion graph of unlabelled Atlastin acceptor proteoliposomes fused with equimolar fluorescently labelled Atlastin donor proteoliposomes. NBD fluorescence was measured at 1 min intervals and detergent was added at 60 min to determine maximum fluorescence. Inset histogram shows the extent of fusion at 60 min. **b,** Surface density concentration-dependence. Surface density is represented as a protein:lipid molar ratio (shown in parentheses). **c,** Kinetic analysis of Atlastin fusion and proteoliposome size change. A single fusion reaction was analysed for fusion by measuring NBD fluorescence and average population size by dynamic light scattering (DLS). **d,** Correlation of size change and fusion. The kinetic data from **c** are plotted as a ratio of fusion percentage versus size.

relies on the action of cytosolic protein components[2]. Although reticulons are the major players in tubularization of the ER[4], no obvious candidates responsible for mediating homotypic fusion of ER membranes have been identified[23].

Our combined *in vivo* and *in vitro* analyses provide strong evidence that Atlastin is the vital GTPase required for homotypic fusion of ER membranes. Atlastin resides in the ER—a localization consistent with a role in mediating homotypic interactions between ER membranes. In response to loss of *atlastin*, the ER network becomes fragmented but no obvious transport impairment was observed, supporting a function in the maintenance of ER integrity rather than in secretory traffic. Nevertheless, subtle transport defects after fragmentation cannot be completely ruled out. Reduced membrane traffic in *Drosophila* results in cell growth defects[24], indicating that minor transport impairment secondary to ER disorganization may explain the small size of Atlastin mutant cells and individuals. Atlastin is capable of homo-oligomerization and self-association can occur within the same membrane as well as between opposing membranes. This property leads to the formation of *trans*-complexes that tether adjacent ER membranes. *In vivo* overexpression of Atlastin results in the expansion of ER elements, consistent with excessive membrane fusion. In agreement with our interpretation of *in vivo* overexpression results, recombinant Atlastin potently drives membrane fusion *in vitro* in a GTP-dependent manner. Atlastin requires GTPase activity to exert its function because GTPase-deficient Atlastin(K51A) is functionally inactive *in vivo*, fails to tether ER membranes owing to its inability to homo-oligomerize, and does not promote membrane fusion *in vitro*. It is likely that the inability of Atlastin(K51A) to fuse is directly related to its inability to self-associate.
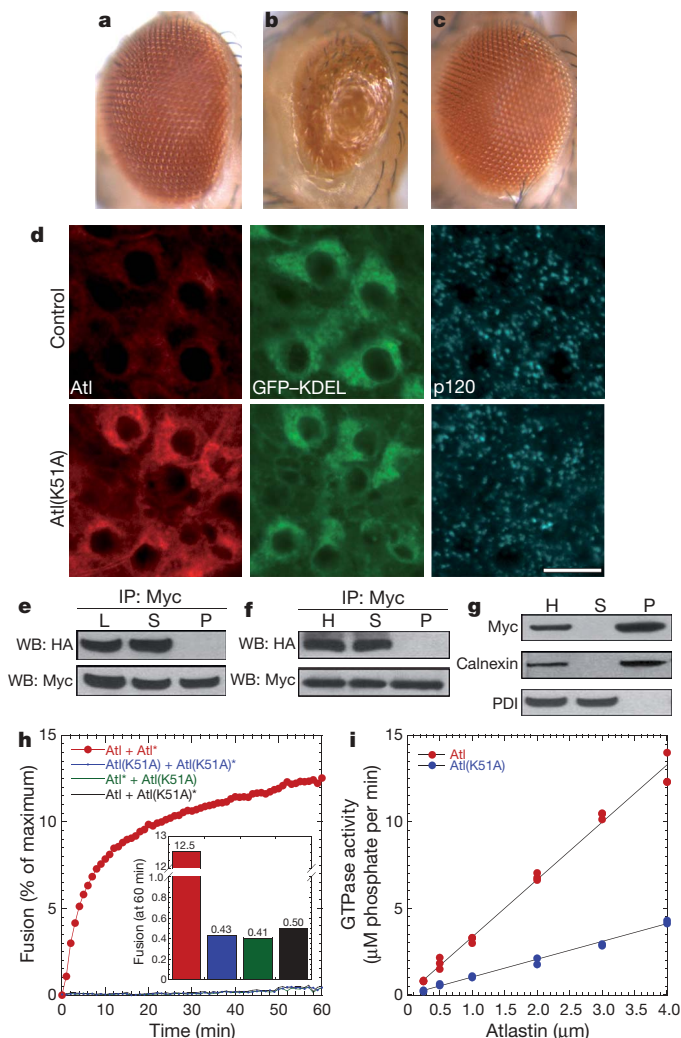
Figure 5 | GTPase-deficient Atlastin is inactivate and unable to mediate membrane tethering and fusion. a, Adult *Drosophila* eye. **b,** Overexpression of Atlastin using GMR-Gal4 generates a small eye. **c,** Eye overexpression of Atlastin(K51A) has no phenotypic consequences. **d,** Simultaneous visualization of GFP–KDEL (green), Golgi p120 (blue) and Atlastin(K51A) (red) of tubulin-Gal4/1;UAS-*atlastin(K51A)-myc*/+ ventral ganglion neurons. Scale bar, 10 μm. **e,** Anti-Myc immunoprecipitates from HeLa cells co-transfected with Atlastin(K51A)–HA and Atlastin(K51A)–Myc are devoid of Atlastin(K51A)–HA. **f,** A vesicle immunoprecipitation assay demonstrates the inability of Atlastin(K51A) to mediate tethering. **g,** Western blot analysis of the soluble and membrane fractions from Atlastin(K51A)-expressing cell homogenates. **h,** Fusion reactions were performed with either donor and/or acceptor proteoliposomes containing Atlastin(K51A). Donor liposomes are indicated with an asterisk. Inset histogram shows extent of fusion at 60 min. **i,** GTPase activity of increasing concentrations of wild-type and mutant Atlastin.

Approximately 50% of the known dominant forms of pure hereditary spastic paraplegia have been linked to ER dysfunction[25]. A properly functioning ER is absolutely required for all cells, given that it is the initial entry point into the secretory pathway for most secreted proteins and membrane proteins localized to the plasma membrane. Disturbance in the function or loss of integrity of the ER could result in a failure of protein folding, glycosylation or transport, leading to ER stress that may ultimately contribute to the pathogenesis of neurodegenerative disorders[26]. Independently of their mechanism, pathological mutations in ATL1 are likely to perturb membrane fusion, with loss of ER integrity suggesting that progressive axonal degeneration in ATL1-linked hereditary spastic paraplegia patients may be the consequence of ER stress engendered by this perturbation.

Our studies have uncovered a requirement for *Drosophila* Atlastin in the homotypic fusion of ER membranes, suggesting that Atlastin represents the GTPase activity thought to be required for this process. Although further studies will be necessary to dissect the structural basis of Atlastin function, the identification of its fusogenic properties lays the foundation for understanding the mechanisms underlying ER biogenesis and maintenance, and may contribute to a better understanding of neuronal degeneration in hereditary spastic paraplegia.

METHODS SUMMARY

Fly culture and transgenesis were performed following standard procedures. Embryos were fixed in 4% formaldehyde in PEM buffer for 20 min before immunolabelling. Larvae and brains were dissected in PBS and fixed in 4% paraformaldehyde. HeLa cell transfection and immunoprecipitation experiments were carried out using standard protocols. For co-immunoprecipitation experiments, Atlastin expression plasmids were simultaneously transfected into HeLa cells. For the vesicle assay, cells were transfected separately, resuspended, homogenized and sonicated. The supernatants containing vesiculated membranes were recovered by centrifugation. Vesiculated membranes were mixed, incubated at 30 °C, and immunoprecipitated. FLIP experiments were performed as described with minor modifications[11]. Pre-embedded gold labelling of Atlastin in S2 cells was done according to the protocol described previously[27]. *Drosophila* brains and larvae were fixed in 4% paraformaldehyde and 2% glutaraldehyde, embedded in EPON, and EM images acquired from thin sections. Recombinant protein was expressed in *Escherichia coli* and purified using a GST-tag. *atlastin* and *atlastin(K51A)* were subcloned into a modified pGEX4T3 vector (GE Healthcare). GST-fusion proteins was expressed in BL21(DE3) (Stratagene) and purified in 0.1% ANAPOE X-100. Unilamellar (100 nm) liposomes were made by extrusion (Avestin Inc.), and purified protein was reconstituted into liposomes by detergent-assisted insertion[28]. Fusion experiments were similar to those previously described[21] with minor modifications[21]. Equimolar amounts of donor and acceptor Atlastin proteoliposomes were mixed, and Mg^{2+} and GTP were added to stimulate fusion. Liposome size was determined by dynamic light scattering (Zen 3600 Zetasizer, Malvern Instruments). GTPase activity of recombinant protein in detergent was measured using an Enzchek phosphate assay kit (Invitrogen).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Anderson, D. J. & Hetzer, M. W. Nuclear envelope formation by chromatin-mediated reorganization of the endoplasmic reticulum. *Nature Cell Biol.* **9**, 1160–1166 (2007).
2. Dreier, L. & Rapoport, T. A. *In vitro* formation of the endoplasmic reticulum occurs independently of microtubules by a controlled fusion reaction. *J. Cell Biol.* **148**, 883–898 (2000).
3. Vedrenne, C. & Hauri, H. P. Morphogenesis of the endoplasmic reticulum: beyond active membrane expansion. *Traffic* **7**, 639–646 (2006).
4. Voeltz, G. K., Prinz, W. A., Shibata, Y., Rist, J. M. & Rapoport, T. A. A class of membrane proteins shaping the tubular endoplasmic reticulum. *Cell* **124**, 573–586 (2006).
5. Zhao, X. *et al.* Mutations in a newly identified GTPase gene cause autosomal dominant hereditary spastic paraplegia. *Nature Genet.* **29**, 326–331 (2001).
6. Sanderson, C. M. *et al.* Spastin and atlastin, two proteins mutated in autosomal-dominant hereditary spastic paraplegia, are binding partners. *Hum. Mol. Genet.* **15**, 307–318 (2006).
7. Zhu, P. P. *et al.* Cellular localization, oligomerization, and membrane association of the hereditary spastic paraplegia 3A (SPG3A) protein atlastin. *J. Biol. Chem.* **278**, 49063–49071 (2003).
8. Lee, Y. *et al.* Loss of spastic paraplegia gene atlastin induces age-dependent death of dopaminergic neurons in *Drosophila. Neurobiol. Aging* **29**, 84–94 (2006).
9. Rismanchi, N., Soderblom, C., Stadler, J., Zhu, P. P. & Blackstone, C. Atlastin GTPases are required for Golgi apparatus and ER morphogenesis. *Hum. Mol. Genet.* **17**, 1591–1604 (2008).
10. Wakefield, S. & Tear, G. The *Drosophila* reticulon, Rtnl-1, has multiple differentially expressed isoforms that are associated with a sub-compartment of the endoplasmic reticulum. *Cell. Mol. Life Sci.* **63**, 2027–2038 (2006).
11. Snapp, E. L., Iida, T., Frescas, D., Lippincott-Schwartz, J. & Lilly, M. A. The fusome mediates intercellular endoplasmic reticulum connectivity in *Drosophila* ovarian cysts. *Mol. Biol. Cell* **15**, 4512–4521 (2004).
12. Stanley, H., Botas, J. & Malhotra, V. The mechanism of Golgi segregation during mitosis is cell type-specific. *Proc. Natl Acad. Sci. USA* **94**, 14467–14470 (1997).
13. Kondylis, V., Goulding, S. E., Dunne, J. C. & Rabouille, C. Biogenesis of Golgi stacks in imaginal discs of *Drosophila melanogaster. Mol. Biol. Cell* **12**, 2308–2327 (2001).

14. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401–415 (1993).
15. Lippincott-Schwartz, J., Snapp, E. & Kenworthy, A. Studying protein dynamics in living cells. *Nature Rev. Mol. Cell Biol.* 2, 444–456 (2001).
16. Parkes, T. L. *et al.* Extension of *Drosophila* lifespan by overexpression of human SOD1 in motorneurons. *Nature Genet.* 19, 171–174 (1998).
17. Ward, T. H., Polishchuk, R. S., Caplan, S., Hirschberg, K. & Lippincott-Schwartz, J. Maintenance of Golgi structure and function depends on the integrity of ER export. *J. Cell Biol.* 155, 557–570 (2001).
18. Ishihara, N., Eura, Y. & Mihara, K. Mitofusin 1 and 2 play distinct roles in mitochondrial fusion reactions via GTPase activity. *J. Cell Sci.* 117, 6535–6546 (2004).
19. Rigaud, J. L. & Levy, D. Reconstitution of membrane proteins into liposomes. *Methods Enzymol.* 372, 65–86 (2003).
20. Scott, B. L. *et al.* Liposome fusion assay to monitor intracellular membrane fusion machines. *Methods Enzymol.* 372, 274–300 (2003).
21. Weber, T. *et al.* SNAREpins: minimal machinery for membrane fusion. *Cell* 92, 759–772 (1998).
22. Praefcke, G. J. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature Rev. Mol. Cell Biol.* 5, 133–147 (2004).
23. Borgese, N., Francolini, M. & Snapp, E. Endoplasmic reticulum architecture: structures in flux. *Curr. Opin. Cell Biol.* 18, 358–364 (2006).
24. Lee, S. & Cooley, L. Jagunal is required for reorganizing the endoplasmic reticulum during *Drosophila* oogenesis. *J. Cell Biol.* 176, 941–952 (2007).
25. Depienne, C., Stevanin, G., Brice, A. & Durr, A. Hereditary spastic paraplegias: an update. *Curr. Opin. Neurol.* 20, 674–680 (2007).
26. Lindholm, D., Wootz, H. & Korhonen, L. ER stress and neurodegenerative diseases. *Cell Death Differ.* 13, 385–392 (2006).
27. Polishchuk, R. S. & Mironov, A. A. Correlative video light/electron microscopy. *Curr. Protoc. Cell Biol.* Chapter 4, Unit 4.8 (2001).
28. Rigaud, J. L., Pitard, B. & Levy, D. Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *Biochim. Biophys. Acta* 1231, 223–246 (1995).

Author Contributions G.O. and D.P. contributed to the experimental design and carried out all *Drosophila* work. D.P. developed and performed the vesicle immunoprecipitation assay. J.T. performed HeLa cell transfection and immunoprecipitation experiments. S.L., T.J.M. and J.E.F. carried out liposome production, analyses and *in vitro* fusion. M.M. and A.E. conducted electron microscopy experiments. A.M. contributed to the experimental design and discussions. J.A.M. designed and supervised liposome production, analyses and *in vitro* fusion and contributed to writing the manuscript. A.D. conceived and designed the study, supervised *Drosophila* and cell culture experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.D. (daga@unipd.it) or J.A.M. (mcnew@rice.edu).

METHODS

Molecular biology. Total RNA from adult *Drosophila* heads was purified with Trizol reagent (Invitrogen). The *atlastin* full-length complementary DNA was obtained by RT–PCR performed on total head RNA. The *atlastin* cDNA was subcloned in the pUAST vector in-frame with a Myc tag. The UAS-*atl*-RNAi construct was made by introducing the entire *atlastin* cDNA lacking the start codon in a Gateway-modified SympUAST vector[29] amplified using the following primers: 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGGCGGATCG GCAGTGCAGGT-3′ and 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTT CATGACCGCTTCACCTTGCCA-3′. Mutagenesis to generate the Lys51Ala mutant was performed with the QuickChange kit (Stratagene).

***Drosophila* genetics.** Fly culture and transgenesis were performed using standard procedures. Several transgenic lines for each construct were generated and tested. *Drosophila* strains used: elav-Gal4; D42-Gal4, GMR-Gal4, tubulin-Gal4, nanos-Gal4:VP16, UAS-mCD8-GFP (Bloomington); MHC-Gal4, Mef2-Gal4; armadillo-Gal4; UAS-αCOP-RNAi, UAS-Sar1-RNAi (VDRC); pUASp:Lys-GFP-KDEL, pUASp:GalT-GFP and Sar1-GFP. Control genotypes varied depending on individual experiments, but always included promoter-Gal4/+ and UAS-transgene/+ individuals.

Immunohistochemistry, HeLa cell transfection, immunoprecipitation and antibodies. Freshly collected embryos were dechorionated, devitellinized, and fixed in 4% formaldehyde in PEM buffer for 20 min before immunolabelling. Third-instar larvae and brains were dissected in PBS and fixed in 4% para-formaldehyde in PBS for 15 min. HeLa cell transfection and immunoprecipitation experiments were carried out using standard protocols.

Quantification of the mean fluorescence intensity of mCD8-GFP at the neuromuscular junction was achieved by normalizing it to the intensity of HRP. For each genotype, eight neuromuscular junctions innervating muscle 4, segment A2, were imaged and statistically analysed. *P* values reported in this study are two-tailed values and derived from a Student's *t*-test, assuming unequal variances.

Confocal images were acquired through ×40 or ×60 CFI Plan Apochromat Nikon objectives with a Nikon C1 confocal microscope and analysed using the NIS Elements software (Nikon). Figure panels were assembled using Adobe Photoshop CS2.

A synthetic peptide (VGGGAASYRSQTSVNASNGKVKRS) was used to immunize guinea pigs for the production of anti-Atlastin antibodies. The following antibodies were used: mouse anti-Myc (1:500, Cell Signaling), rabbit anti-Myc (1:200, Santa Cruz Biotechnology), mouse anti-HA (1:1,000, Cell Signaling), guinea pig anti-Atlastin (1:500), rat anti-BiP (1:50, Babraham), mouse anti-p120 (1:600, Calbiochem), mouse anti-GFP (1:500, Roche), mouse anti-Dlg (1:100, DSHB), mouse anti-PDI (1:500, Stressgen), rabbit anti-calnexin (1:1,000, Millipore). Secondary antibodies for immunofluorescence (Cy5 and Cy3 conjugates from Jackson laboratories, Alexa Fluor 488 conjugates from Invitrogen) were used at 1:1,000. Anti-mouse, anti-guinea pig and anti-rabbit HRP conjugates from DACO were used at 1:10,000.

Immuno-isolation of membrane vesicles containing Atlastin and membrane fractionation. Expression plasmids for HA-tagged or Myc-tagged Atlastin were separately transfected into HeLa cells. Cells were suspended in homogenization buffer (10 mM HEPES-KOH buffer, pH 7.4, containing 0.22 M mannitol, 0.07 M sucrose and protease inhibitors), and homogenized using a syringe with a 26-gauge needle. Homogenates were sonicated and the supernatants containing vesiculated membranes recovered by centrifugation at 4,000*g* for 5 min to remove unbroken organelles. Vesiculated membranes were mixed, and incubated at 30 °C for 1 h. After incubation, immunoprecipitation of Atlastin-containing vesicles was performed in the absence of detergent following standard procedures. The remaining supernatants containing vesiculated membranes were centrifuged at 120,000*g* for 60 min to separate a membrane fraction (pellet) and a soluble fraction (supernatant). Supernatants and pellets derived from the immunoprecipitation and 120,000*g* centrifugation were analysed by western blotting.

Fluorescence loss in photobleaching. FLIP experiments were performed as described with minor modifications[11]. Experimental larvae expressing UAS-GFP-KDEL were dissected in Ca^{2+}-free HL3 and analysed using a Zeiss LSM 510 confocal microscope with a 0.3 numerical aperture (NA) ×40 water immersion objective. Two different ROI for each genotype distributed along muscle 6 or 7 in the abdominal segment 4 were selected and bleached by 20 iterations, at 100% laser power, followed by three scanning images every 15 s. The bleaching protocols were repeated for 1 h. The area of ROI in *atl* RNAi muscle was approximately 60% of the ROI area in control muscle to reflect the smaller size of *atl* RNAi larvae. To create fluorescence recovery curves, fluorescence intensities were transformed into a 0–100% scale and were plotted using Excel software. FLIP experiments were repeated at least three times.

Electron microscopy. Pre-embedded gold labelling of Atlastin was done according to the protocol described previously[27]. In brief, S2 cells were fixed in 4%

paraformaldehyde and 1% glutaraldehyde, washed, incubated with the anti-Atlastin antibody (1:100) overnight and then with Nanogold conjugated Fab fragments of the secondary antibodies (Nanoprobes) for 2 h. The Nanogold particles were developed using the Gold-enhance kit. Labelled cells were dehydrated, embedded in Epon, and sectioned. *Drosophila* brains and larvae were fixed in 4% paraformaldehyde and 2% glutaraldehyde and embedded as described earlier. EM images were acquired from thin sections under a Philips Tecnai-12 electron microscope using an ULTRA VIEW CCD digital camera (Soft Imaging Systems GmbH). Quantification of anti-Atlastin labelling was done by counting gold particles in 65 cells. EM images of individual neurons for the measurement of the length of ER profiles were collected from three brains for each genotype. In total, 18 *atl* RNAi, 20 *atl*^l and 17 control neurons were analysed. In the rescue experiment a total of 26 neurons expressing wild-type Atlastin and 23 neurons expressing Atlastin(K51A) were analysed. Quantitative analyses were performed with AnalySIS software (Soft Imaging Systems GmbH).

Recombinant Atlastin production. The double-stranded oligonucleotides XhoI-His$_8$-Top (5′-TCGAGCATCATCATCATCATCATCATCACTGAC-3′) and His$_8$-bottom-EagI (5′-GGCCGTCAGTGATGATGATGATGATGATGAT GC-3′) were ligated into pGEX-4T-3 cut with XhoI and NotI to generate pJM680 (pGEX-4T-3 with a carboxy-terminal octa-histidine tag). *Drosophila melanogaster atlastin* was PCR amplified from pGEX-*atlastin* using the oligonucleotides BamHI-Dm_Atl (5′-CCGGATCCATGGGCGGATCGGC-3′) and Dm_Atl-XhoI(w/o stop) (5′-ATCTCGAGTGACCGCTTCACC-3′), digested with BamHI and XhoI, and ligated into pJM680 cut with the same enzymes to generate pJM681. An internal fragment containing the Lys51Ala mutation was digested from pGEX-*atlastin(K51A)* using PpuMI and SacI and ligated into pJM681 cut with the same enzymes to generate pJM694. pJM681 and pJM694 were transformed into *E. coli* BL21(DE3) (Stratagene) and cultures were grown in Luria-Bertani broth with 100 μg ml^{-1} ampicillin. Five millilitres of pre-precultures were grown overnight at 37 °C. Fifty millilitres of precultures were inoculated with the pre-precultures and grown for ~6 h at 37 °C. Four litre cultures were inoculated to a density (D_{600}) of 0.03 with cells collected from the precultures and grown at 37 °C until a D_{600} of 0.4. The temperature was lowered to 16 °C and cells were induced with 0.1 mM IPTG when the D_{600} reached 0.6 and grown overnight at 16 °C. Cells were collected, washed with 25 mM HEPES, pH 7.4, and 200 mM KCl, and lysed in 25 mM HEPES, pH 7.4, 200 mM KCl, 10% glycerol, 2 mM 2-mercaptoethanol, 4% Triton X-100, 2 mM EDTA and complete protease inhibitors (Roche). Extracts were centrifuged for 30 min at 70,000*g*. Extracts were incubated with 70 mg of swelled GSH-agarose beads for 3 h at 4 °C. Beads were washed with 25 mM HEPES, pH 7.4, 100 mM KCl, 10% glycerol, 2 mM 2-mercaptoethanol, 1% Triton X-100, and 1 mM EDTA, followed by 25 mM HEPES, pH 7.4, 100 mM KCl, 10% glycerol, 2 mM 2-mercaptoethanol, 0.1% Triton X-100, and 1 mM EDTA. Protein was eluted from GSH-agarose with 25 mM HEPES, pH 7.4, 100 mM KCl, 10% glycerol, 2 mM 2-mercaptoethanol, 0.1% Triton X-100, 1 mM EDTA, and 10 mM glutathione.

Atlastin reconstitution and proteoliposome production. Liposome production. Atlastin in 0.1% Triton X-100 (Fisher) was reconstituted into pre-formed 100 nm liposomes. Liposomes were made by extrusion method as follows[30]: unlabelled lipid mixes (POPC:DOPS, 85:15 molar ratio) and labelled lipid mixes (POPC:DOPS:Rho-DPPE:NBD-DPPE, 83:15:1.5:1.5 molar ratio) with trace amounts of ^{3}H-1-palmitoyl 2-palmitoylphosphatidylethanolamine (DPPE) in chloroform were dried under a stream of N_2 gas followed by further drying in a vacuum for 30 min. The lipid films were the resuspended in buffer A100 (25 mM HEPES, pH 7.4, 100 mM KCl, 10% glycerol, 2 mM β-mercaptoethanol, 1 mM EDTA) to a final total lipid concentration of about 10 mM. Large unilamellar vesicles (LUVs) were formed by ten freeze-thaw cycles in liquid N_2 and room temperature water. Uniform-sized LUVs were formed by extrusion through polycarbonate filters with 100-nm pore size (Avanti Polar Lipids).

Total lipid concentration of liposomes was determined by scintillation counting and liposomes size was confirmed by DLS. Atlastin was reconstituted into both labelled and unlabelled 100 nm liposomes by detergent assisted insertion[30] as follows: Atlastin in 0.1% Triton X-100 was mixed with preformed liposomes at an effective detergent to lipid ratio of ~1. The effective detergent to lipid ratio (R_{eff}) was determined by the equation $R_{eff} = (D_{total} - D_{water})/[lipid]$, in which D_{total} is the total detergent concentration and D_{water} is the monomeric detergent concentration (0.18 mM) in the presence of detergent[19]. Protein and lipid were allowed to mix for 1 h at 4 °C. Detergent was removed by adding BioBeads SM-2 adsorbant beads (BioRad). Insoluble protein aggregates were pelleted by centrifugation of the samples in an Eppendorf microcentrifuge (10 min, 16,000*g*). Final lipid and protein concentration were determined by liquid scintillation counting and amido black protein assay[31], respectively.

***In vitro* fusion assays.** Fusion assays were on the basis of methods previously described[21] with the following modifications: labelled and unlabelled populations

of Atlastin proteoliposomes (0.3 mM lipid each) were mixed in the presence of 5 mM Mg^{2+} and brought to total volume of 47.5 μl with A100 buffer in each well of a 96-well Fluoronunc polysorp plate (Nunc). Samples in the plate were incubated at 37 °C in a fluorescent plate reader (Infinite M200, TECAN) for 5 min. To each well, 2.5 μl 20 mM GTP or A100 buffer (control) was added and NBD fluorescence was measured (excitation 460 nm, emission 538 nm) at 1 min intervals for 60 min with 1 s shaking after every read. To determine the absolute NBD fluorescence, 10 μl of 2.5% (w/v) n-dodecylmaltoside was added at 60 min. Data was normalized to the percentage total fluorescence. For Fig. 4a, the extent of fusion at 60 min in shown as a histogram inset. The absolute values are: GTP and Mg^{2+} 11.3%, GTP 0.26%, Mg^{2+} 0.23%, GTPγS 0.21% and buffer 0.23%. For Fig. 5h, the extent of fusion at 60 min in shown as a histogram inset. The absolute values (with asterisk representing the donor) are: Atl + Atl* 12.5%, Atl(K51A) + Atl(K51A)* 0.43%, Atl* + Atl(K51A) 0.41%, and Atl + Atl(K51A)* 0.50%. All fusion reactions had a 1:200 Atlastin:lipid ratio except as shown in Fig. 4b.

Proteoliposome population size determination by DLS. All the DLS data were acquired using a Malvern Zetasizer Nano ZS and ZEN0040 cuvettes. The scattering of light was recorded at 173 °C (back scattering). The light intensity and the measurement time were automatically set by the instrument software. To measure the dynamic change of liposome size, the experiments were performed at 37 °C with 300 μM liposome in A100 buffer (100 mM KCl, 25 mM HEPES, 10% glycerol, 1 mM EDTA). Different nucleotides, Mg^{2+} or EDTA were added at the indicated time. To measure the size change of the liposomes during fusion, three identical fusion assays were used. One fusion assay was performed as mentioned earlier to acquire fusion data. Samples from two other fusion assays were taken out at various time points and titrated into EDTA buffer (100 mM KCl, 25 mM HEPES, 10% glycerol, 5 mM EDTA) to stop the fusion reaction.

These samples were analysed three times by DLS and the average of these three measurements was used to represent the size of the liposomes. For Fig. 4c, the population increased in size from 136 nm at time 0 to 390 nm at 60 min.

GTPase activity measurements. GTPase activity was determined by measuring the release of inorganic phosphate from GTP using the EnzChek Phosphate Assay Kit (Molecular Probes) according to the manufacturer's instructions. In brief, recombinant Atlastin in detergent was mixed in a 250 μl reaction volume with $1 \, U \, ml^{-1}$ purine nucleoside phosphorylase (PNP), 200 μM 2-amino-6-mercapto-7-methylpurine riboside (MESG) and 400 μM GTP in a transparent 96-well plate (Nunc), and incubated at room temperature for 10 min. The plate was then warmed to 37 °C in a microplate reader (Tecan, Infinite M200) and 5 mM Mg^{2+} was added to start the reaction. PNP catalyses the conversion of MESG to ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine in the presence of inorganic phosphate, resulting in a spectrophotometric shift in absorbance from 330 to 360 nm. The absorbance increase at 360 nm was measured in real-time every 30 s over 30 min. Absorbance was normalized to phosphate standards and initial rates were calculated. The data in Fig. 4d represents two separate protein preparations of wild-type and mutant Atlastin.

29. Giordano, E., Rendina, R., Peluso, I. & Furia, M. RNAi triggered by symmetrically transcribed transgenes in *Drosophila melanogaster*. *Genetics* **160**, 637–648 (2002).

30. Schaub, J. R., Lu, X., Doneske, B., Shin, Y. K. & McNew, J. A. Hemifusion arrest by complexin is relieved by Ca^{2+}-synaptotagmin I. *Nature Struct. Mol. Biol.* **13**, 748–750 (2006).

31. Schaffner, W. & Weissmann, C. A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal. Biochem.* **56**, 502–514 (1973).

ARTICLES

Specific pathways prevent duplication-mediated genome rearrangements

Christopher D. Putnam[1], Tikvah K. Hayes[1] & Richard D. Kolodner[1]

We have investigated the ability of different regions of the left arm of *Saccharomyces cerevisiae* chromosome V to participate in the formation of gross chromosomal rearrangements (GCRs). We found that the 4.2-kilobase *HXT13-DSF1* region sharing divergent homology with chromosomes IV, X and XIV, similar to mammalian segmental duplications, was 'at risk' for participating in duplication-mediated GCRs generated by homologous recombination. Numerous genes and pathways, including *SGS1, TOP3, RMI1, SRS2, RAD6, SLX1, SLX4, SLX5, MSH2, MSH6, RAD10* and the DNA replication stress checkpoint requiring *MRC1* and *TOF1*, were highly specific for suppressing these GCRs compared to GCRs mediated by single-copy sequences. These results indicate that the mechanisms for formation and suppression of rearrangements occurring in regions containing at-risk sequences differ from those occurring in regions of single-copy sequence. This explains how extensive genome instability is prevented in eukaryotic cells whose genomes contain numerous divergent repeated sequences.

The importance of maintaining the stability of the genome is revealed by the numerous genetic diseases caused by inherited and *de novo* mutations ranging from base changes to genome rearrangements[1,2]. In addition, many cancers are associated with ongoing genome instability and the continued accumulation of mutations and genome rearrangements[3–7]. Despite the problems introduced by genome instability, the human genome contains many features prone to be unstable, including microsatellite repeats, minisatellite repeats, triplet repeats, short separated repeats, mirror repeats, inverted repeats, and dispersed repetitive elements such as retroviral elements, short interspersed repeated elements (SINEs), long interspersed repeated elements (LINEs), segmental duplications and regions of copy number variation (CNVs)[8,9]. Dispersed repetitive elements can underlie chromosomal rearrangements through non-allelic homologous recombination (HR) between elements at non-homologous chromosomal locations. Alu elements, for example, cause HR-mediated deletions, duplications and chromosomal translocations implicated in over 15 inherited diseases as well as rearrangements leading to cancer[10]. Similarly, more than 20 human diseases are caused by rearrangements mediated by non-allelic HR between segmental duplications[11]. Given the large numbers of repeated regions in the genome, it is surprising that the genome is as stable as it is.

Some types of at-risk sequences have been characterized in *Saccharomyces cerevisiae*[9]. Engineered duplications are targets of ectopic recombination, leading to both gene conversion and chromosomal rearrangements[12]. Similarly, Ty transposons, which are dispersed, repeated sequences, can recombine to produce genome rearrangements[13], and inverted copies of Ty transposons can cause double-strand breaks (DSBs) during replication, resulting in genome rearrangements[14]. Consistent with this, at-risk sequences seem to be selected against[15]; however, the human genome still retains many such sequences. Although at-risk sequences can induce genome instability, little is known about how such genome rearrangements are prevented and whether there are specific pathways that prevent such sequences from causing genome rearrangements.

We have described assays for measuring the rate of accumulating GCRs[16]. This assay system detects GCRs that occur in natural DNA

sequences and does not depend on the introduction of artificial DNA sequences or the artificial induction of DSBs. Here, we applied this system to compare the rates and properties of GCRs in a chromosomal region devoid of at-risk sequences with that of a region of the genome containing a sequence homeologous to ectopic regions of the genome reminiscent of segmental duplications.

Duplications alter the GCR spectrum and rate

We placed a *CAN1/URA3* cassette in different locations on the non-essential left end of chromosome V to select for canavanine (Can)- and 5-fluoroorotate (5-FOA)-resistant GCRs similar to our original GCR assay[16] (Fig. 1a). GCRs, but not co-mutation or interstitial co-deletion of *CAN1* and *URA3*, dominated the Can[r] 5-FOA[r] products as demonstrated by frequent loss of a telomeric hygromycin-resistance marker (Supplementary Table 1), similar to the original GCR assay[17]. Overall, the GCR rates increased approximately linearly with the size of the chromosome V breakpoint region except for the *yel072w::CAN1/URA3* assay, which had a higher rate than predicted based on the breakpoint region length (Table 1). *YEL072W* is telomeric to the *HXT13-DSF1* region, which shares ~4.2 kilobases (kb) of imperfect homology with chromosome XIV and ~2 kb of imperfect homology with nearly identical regions of chromosomes IV and X (Fig. 1b), similar to mammalian segmental duplications[18]. Deletion of *HXT13-DSF1* eliminated the duplication-associated GCR rate increase (Table 1). Homology-driven monocentric t(V;XIV) and t(V;IV or X) translocations accounted for 90% of the GCRs even though the *HXT13-DSF1* region accounts for 13% of the breakpoint region (Fig. 2a). Sequencing of 20 t(V;XIV) junctions only revealed translocation breakpoints in the *HXT13-DSF1* homology regions (Supplementary Fig. 1a)[17]. Array comparative genomic hybridization (aCGH) demonstrated that the target chromosomes were duplicated from the homology to the telomere (Fig. 1c), indicating that an intact copy of the target chromosomes was maintained; this was confirmed by polymerase chain reaction (PCR) amplification of the native *HXT13-DSF1* related junctions on the target chromosome (data not shown). Overall, the homology-driven GCRs were consistent with break-induced replication (BIR) or related mechanisms[19,20].

[1]Ludwig Institute for Cancer Research, Departments of Medicine and Cellular and Molecular Medicine and Cancer Center, University of California School of Medicine, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0669, USA.
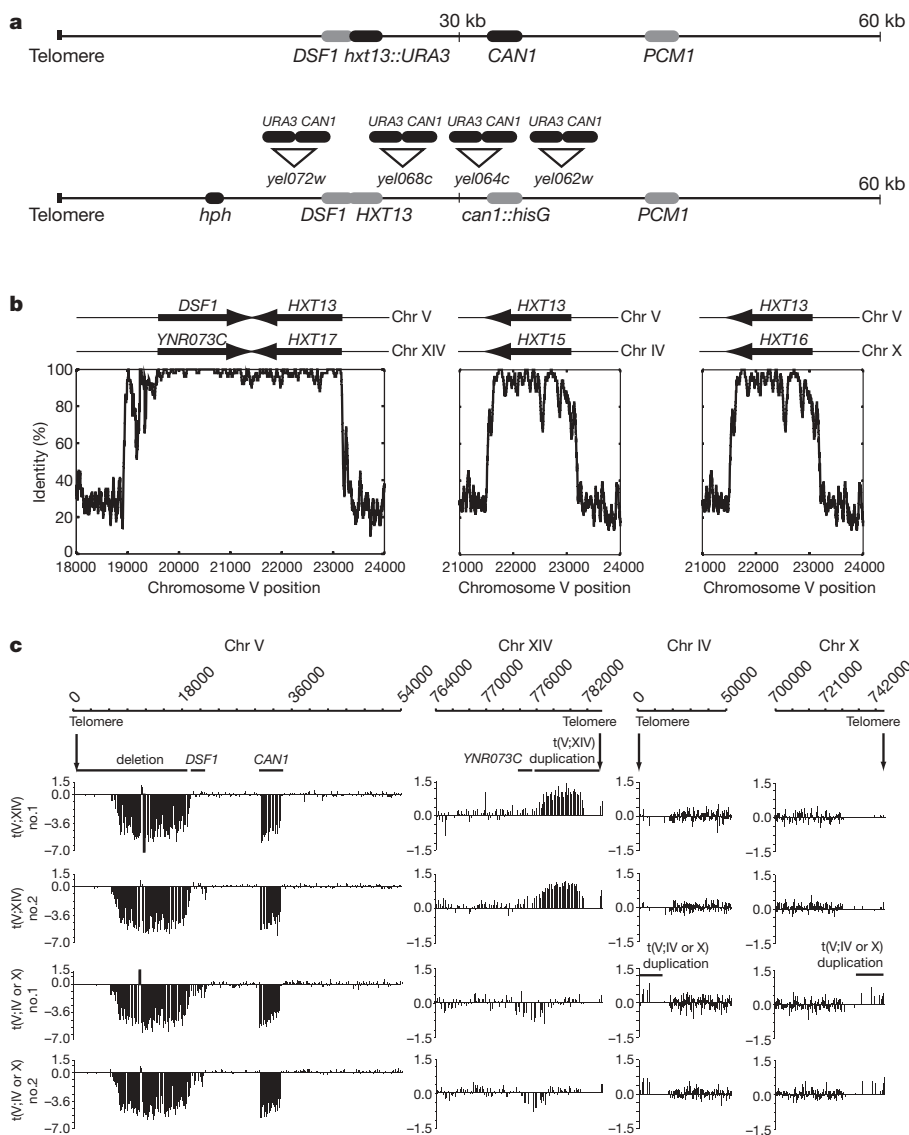
Figure 1 | New assays for evaluating the genes that suppress the accumulation of GCRs. a, The standard chromosome V GCR assay (top) contains *URA3* integrated at *HXT13* and selects for GCRs with chromosome V breakpoints located between *CAN1* and the essential *PCM1* gene. The modified GCR assays (bottom) have a *CAN1/URA3* cassette inserted into *YEL062W*, *YEL064C*, *YEL068C* and *YEL072W* in a strain with *ura3-52* and *can1::hisG* mutations and a telomeric hygromycin-resistance marker (*hph*). **b,** The average per cent identity in 50-bp windows with the *HXT13-DSF1* region with regions of chromosomes XIV, X and IV is plotted against the chromosome V position. **c,** aCGH data (log$_2$ of the fluorescence ratio of individual GCR isolates to wild type) indicate that the region from the chromosome V homologies to the target chromosome telomere was duplicated. The two t(V;XIV) fusions lost unique chromosome V signals telomeric to the *HXT13-DSF1* region (chromosome V 1–19500) and *CAN1* from the *CAN1/URA3* cassette (chromosome V 31694–33466). Increased signals were observed with all probes for chromosome XIV telomeric to *YNR073C* (chromosome XIV 776300–787000). The two t(V;IV or X) fusions had chromosome V signals similar to the t(V;XIV) fusions and essentially unchanged chromosome XIV signals, except for a subtle loss of signal in the *DSF1* and *YNR073C* regions (chromosome V 19589–21097; chromosome XIV 774792–776300), consistent with loss of cross hybridization of *DSF1* DNA to probes for *DSF1*-like genes. Increased fluorescence of the left arm of chromosome IV and the right arm of chromosome X demonstrated amplification and cross hybridization between these almost identical regions, despite the scarcity of probes. The aCGH data revealed no other significant copy number changes, except for the region indicating loss of *URA3* from the *CAN1/URA3* cassette (data not shown).

Genotype affects the impact of the duplication

In the standard GCR assay, deletion of *MRE11* or *RAD27* caused ~600–1,000-fold increased GCR rates[16] and caused similar rate increases in strains where the *CAN1/URA3* cassette was centromeric to *HXT13-DSF1* (Table 1). When the cassette was telomeric to the duplication, the mutations only caused a modest increase in GCR rate relative to the *yel068c::CAN1/URA3* assay: fivefold for *rad27Δ* and threefold for *mre11Δ*. The GCRs in the *rad27Δ yel072w::CAN1/URA3* background were a mix of 33 duplication-mediated and 16 single-copy-sequence-mediated products (Fig. 2a). Like the wild-type strain that had 56 duplication-mediated and 6 single-copy-sequence-mediated GCRs, the ratios of products were similar to the fold increase in rate caused by the duplication. In contrast, the *mre11Δ yel072w::CAN1/URA3* GCRs were dominated by duplication-mediated rearrangements (56:1). This suggests that an *mre11Δ* mutation alters the mechanism underlying GCRs such as by decreasing telomere maintenance[21], resulting in increased degradation of chromosome ends, which would preferentially target telomeric duplicated sequences.

Deletion of *SGS1*, encoding a RecQ-family helicase, caused a moderate increase in the rate of GCRs in assays with the *CAN1/URA3* cassette centromeric to the duplication similar to the standard GCR assay, but caused a marked increase in the *yel072w::CAN1/URA3* GCR rate that depended on the *HXT13-DSF1* duplication (Table 1). The ratio of duplication-mediated to single-copy-sequence-mediated

GCRs in the *sgs1Δ* mutant, 35:5 (Fig. 2a), was not as high as might be predicted from the 115-fold increase in GCR rates in the *yel072w::CAN1/URA3* assay versus the *yel068c::CAN1/URA3* assay. Sequencing of 25 *sgs1Δ* t(V;XIV) breakpoint junctions revealed 21 t(V;XIV) and 4 complex translocations (Supplementary Fig. 1b). Three complex breakpoints resulted from t(V;XIV;V;XIV) translocations, and the fourth was consistent with a t(V;X;XIV) translocation. The complex junctions could be generated by template switching during HR as implicated during BIR in wild-type strains[22] and *CAN1-LYP1-ALP1* translocations in *sgs1Δ* mutants[23], or by formation of multiduplex joint molecules as observed in meiosis[24].

Different HR pathways yield distinct GCR signatures

We next examined the role of the *RAD52* epistasis group genes (Table 2). As in the standard GCR assay[25], the *rad52Δ* mutation increased the GCR rate in the *yel068c::CAN1/URA3* assay where GCRs are formed in single-copy DNA sequences (Table 2). In contrast, the *rad52Δ* mutation modestly decreased the GCR rate in the *yel072w::CAN1/URA3* assay compared to wild type (Table 2) and eliminated the duplication-mediated translocations (Fig. 2b). Deletion of *RAD51* or *RAD59*, which define two distinct *RAD52*-dependent HR pathways[26], had modest effects on the GCR rates in both assays, and non-reciprocal duplication-mediated translocations were observed in both single mutants (Table 2, Fig. 2b and Supplementary Fig. 2), indicating that these rearrangements are

Table 1 | GCR rates for different positions of the *CAN1/URA3* cassette on chromosome V

Assay	Wild type		rad27Δ		mre11Δ		sgs1Δ		Breakpoint region length (kb)†	Wild-type rate/ length (kb⁻¹)
	Strain RDKY no.	Rate*	Strain RDKY no.	Rate*	Strain RDKY no.	Rate*	Strain RDKY no.	Rate*		
Standard GCR‡	3615	3.5×10^{-10} (1)	3630	3.7×10^{-7} (1,057)	3633	2.2×10^{-7} (629)	3813	2.5×10^{-8} (71)	11.6	3.0×10^{-11}
yel062w::CAN1/URA3	6675	1.15 [0.0–5.6] $\times 10^{-10}$ (0.3)	6679	6.87×10^{-7} (2,180)	6680	3.23×10^{-7} (496)	6681	1.19×10^{-8} (34)	9.7	1.2 [0.0–5.7] $\times 10^{-11}$
yel064c::CAN1/URA3	6676	5.09 [2.5–7.7] $\times 10^{-10}$ (1.6)	6682	7.47×10^{-7} (2,371)	6683	2.63×10^{-7} (465)	6684	1.77×10^{-8} (51)	14.5	3.5 [1.7–5.3] $\times 10^{-11}$
yel068c::CAN1/URA3	6677	2.27 [1.3–4.8] $\times 10^{-9}$ (7.2)	6685	5.57×10^{-7} (1,591)	6686	5.75×10^{-7} (1,643)	6687	1.69×10^{-8} (48)	19.2	12 [6.8–25] $\times 10^{-11}$
yel072w::CAN1/URA3	6678	1.97 [1.6–4.3] $\times 10^{-8}$ (56)	6688	2.78×10^{-6} (7,943)	6689	1.52×10^{-6} (4,345)	6690	1.93×10^{-6} (5,515)	31.0	64 [52–140] $\times 10^{-11}$
yel068c::CAN1/URA3 hxt13-dsf1Δ	6872	1.43 [0.0–4.2] $\times 10^{-9}$ (4.1)	6873	3.74×10^{-7} (1,068)	-	ND	6874	4.39×10^{-9} (12)	19.2	7.5 [0.0–22] $\times 10^{-11}$
yel072::CAN1/ URA3 hxt13-dsf1Δ	6875	5.64 [4.1–12] $\times 10^{-9}$ (16)	6876	5.22×10^{-7} (1,492)	-	ND	6877	1.17×10^{-8} (33)	27.7	20 [15–42] $\times 10^{-11}$

* Rate of accumulating Can^r 5-FOA^r progeny. The number in parentheses is the fold increase relative to the standard wild-type rate (3.5×10^{-10}; ref. 16). Numbers in brackets are the 95% confidence interval limits.
† The breakpoint region is defined as the length between the telomeric end of *PCM1* and the telomeric end of *CAN1*.
‡ Rates previously published[16,29].

not strictly dependent on either pathway. t(V;IV or X) translocations were not observed in the *rad59Δ* strain, indicating that efficient recombination with the translocation target that was shorter and had lower sequence identity was *RAD59* dependent. Both the *rad51Δ rad59Δ* double mutant and the *rad51Δ rad59Δ rad52Δ* triple mutant had decreased rates of duplication-mediated GCRs (Table 2). Surprisingly, t(V;XIV) rearrangements were observed in the *rad51Δ rad59Δ* double mutant, unlike the *rad52Δ* single mutant and the *rad51Δ rad59Δ rad52Δ* triple mutant (Fig. 2b). Thus, it seems that at least one additional *RAD52*-dependent, *RAD51*- and *RAD59*-independent HR pathway can promote interchromosomal HR-mediated rearrangements at low rates; this is consistent with a more severe HR defect in a *rad52Δ* single mutant compared to a *rad51Δ rad59Δ* double mutant[27].

Mismatch repair (MMR) proteins[28] and Sgs1 (ref. 29) are predicted to suppress HR between the *HXT13-DSF1* region and the imperfect homologies on chromosomes IV, X and XIV. Elimination of mismatch detection by a *msh2Δ* mutation or impairment by *msh6Δ* or *msh3Δ* mutations specifically increased the GCR rates

in the duplication-containing assay (Table 2). The larger effects of *msh2Δ* and *msh6Δ* relative to *msh3Δ* are consistent with the heteroduplexes formed during duplication-mediated HR, which would contain primarily base–base mispairs and fewer insertion/deletion mispairs. Similar to the effects of *mlh1Δ* in single-stranded annealing assays[30], *mlh1Δ* caused a smaller but significant increase in the rate of duplication-mediated GCRs (Table 2). An *sgs1Δ* mutation caused an increase in duplication-mediated GCRs (Table 1 and Fig. 2), and a *rad52Δ* mutation eliminated this increase (Table 2), indicating that homeologous recombination mediates most of the GCRs in the *yel072w::CAN1/URA3* assay in the *sgs1Δ* mutant. However, *sgs1Δ* caused a higher duplication-mediated GCR rate than *msh2Δ* (Table 2), despite the fact that Sgs1 is downstream of MMR during suppression of homeologous recombination[28]. Deletion of *TOP3* and *RMI1*, which function together with *SGS1* (ref. 31), also caused higher rates of duplication-mediated GCRs than the *msh2Δ* mutation; the increased GCR rates caused by *rmi1Δ* relative to *sgs1Δ* and *top3Δ* suggest that *RMI1* may have *SGS1*- and *TOP3*-independent roles (Table 2). These data, in combination with the synergistic increase in the GCR rate in the *yel068c::CAN1/URA3* assay in the *sgs1Δ rad52Δ* double mutant (Table 2), suggest that *sgs1Δ* as well as *top3Δ* and *rmi1Δ* cause defects in suppression of homeologous recombination and also affect other pathways that suppress duplication-mediated GCRs.

HR-mediated GCRs are *POL32* independent

The t(V;XIV) and t(V;IV or X) translocation products and their dependence on HR genes are consistent with BIR or related mechanisms[19,20]. *POL32*, encoding a DNA polymerase delta subunit, is essential for ectopic BIR induced by HO-mediated DSBs[32], but is not strictly required for allelic BIR[20]. Deletion of *POL32* caused a small increase in the duplication-mediated GCR rate and did not change the rate of t(V;XIV) or t(V;IV or X) translocations (Table 2 and Fig. 2b). The three *pol32Δ* t(V;XIV) translocations analysed by aCGH were non-reciprocal (Supplementary Fig. 3). These results could be explained if previously observed *POL32*-dependent BIR was predominantly *RAD51* dependent[32], in contrast to both *RAD51*-dependent and *RAD51*-independent pathways observed here.

Both the *pol32Δ rad51Δ* and *pol32Δ rad59Δ* double mutants had low levels of duplication-mediated GCRs (Table 2 and Fig. 2b). The *pol32Δ rad51Δ* double mutant had increased GCR rates in both assays, with the duplication causing a modest increase primarily due to accumulation of t(V;IV or X) translocations, consistent with the possibility that *RAD51* is required to suppress GCRs in a *pol32Δ* mutant. In contrast, the *pol32Δ rad59Δ* double mutant had a lower GCR rate than the *rad59Δ* and *pol32Δ* single mutants, and, compared to the
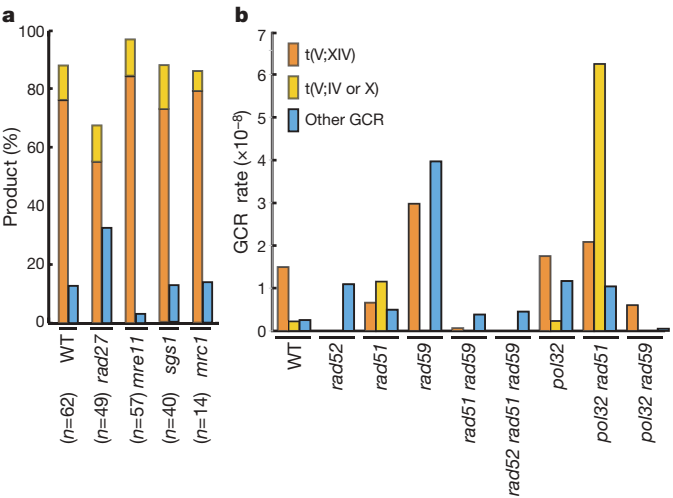
Figure 2 | Summary of the types of GCRs detected in the *HXT13-DSF1*-region-mediated GCR assay. a, Percentage of the different types of GCRs in wild-type, *rad27Δ*, *mre11Δ*, *sgs1Δ* and *mrc1Δ yel072w::CAN1/URA3* strains. The homology-driven GCRs are shown as a stacked bar with t(V;XIV) in orange and t(V;IV or X) in yellow, and single-copy-sequence-mediated GCRs in blue. **b,** Mutations affecting both HR and BIR alter the rates of the formation of t(V;XIV), t(V;IV or X) and other GCRs detected in the *yel072w::CAN1/URA3* assay. Rates for each class of GCR were calculated by multiplying the fraction of each kind of rearrangement by the overall rate.

Table 2 | Effect of homologous and homeologous recombination-defective mutations on GCR rates

Genotype	yel068c:: CAN1/URA3		yel072w:: CAN1/URA3		Ratio†
	Strain	GCR rate*	Strain	GCR rate*	
Wild type	RDKY6677	2.27×10^{-9} (5.1)	RDKY6678	1.97×10^{-8} (56)	8.7
rad52Δ	RDKY6691	1.67×10^{-8} (48)	RDKY6708	1.09×10^{-8} (31)	0.7
rad51Δ	RDKY6692	$<2.63 \times 10^{-10}$ (<0.8)	RDKY6709	2.31×10^{-8} (66)	>88
rad59Δ	RDKY6693	5.85×10^{-9} (17)	RDKY6710	6.94×10^{-8} (198)	11.9
rad51Δ rad59Δ	RDKY6694	2.92×10^{-9} (8.3)	RDKY6711	4.48×10^{-9} (13)	1.5
rad51Δ rad59Δ rad52Δ	RDKY6695	3.84×10^{-9} (11)	RDKY6712	4.53×10^{-9} (13)	1.2
msh2Δ	RDKY6696	1.10×10^{-9} (3.1)	RDKY6713	1.75×10^{-7} (499)	159
msh6Δ	RDKY6697	1.52×10^{-9} (4.4)	RDKY6714	2.10×10^{-7} (599)	138
msh3Δ	RDKY6698	1.42×10^{-9} (4.1)	RDKY6715	3.67×10^{-8} (105)	26
mlh1Δ	RDKY6699	5.80×10^{-10} (1.7)	RDKY6716	3.85×10^{-8} (110)	66
sgs1Δ	RDKY6687	1.69×10^{-8} (48)	RDKY6690	1.93×10^{-6} (5,515)	114
sgs1Δ rad52Δ	RDKY6700	7.75×10^{-8} (222)	RDKY6717	8.07×10^{-8} (231)	1.0
top3Δ	RDKY6701	$<1.64 \times 10^{-9}$ (<4.7)	RDKY6718	2.14×10^{-6} (6,103)	>1,300
rmi1Δ	RDKY6702	1.41×10^{-7} (404)	RDKY6719	1.27×10^{-5} (36,700)	98.3
pol32Δ	RDKY6703	3.41×10^{-9} (9.4)	RDKY6720	3.15×10^{-8} (90.0)	9.24
pol32Δ rad51Δ	RDKY6704	3.00×10^{-8} (86)	RDKY6721	9.37×10^{-8} (268)	3.1
pol32Δ rad59Δ	RDKY6705	5.50×10^{-10} (1.6)	RDKY6722	6.52×10^{-9} (19)	12
pri2-1 (23 deg)	RDKY6706	$<3.93 \times 10^{-10}$ (<1.1)	RDKY6723	7.10×10^{-10} (2.0)	>1.8
pol12-100 (23 deg)	RDKY6707	6.35×10^{-9} (18.1)	RDKY6724	2.31×10^{-8} (66.1)	3.7

* Rate of accumulating Can^r 5-FOA^r progeny. The number in parentheses is the fold increase relative to the standard wild-type rate (3.5×10^{-10}; ref. 16).
† The yel072w::CAN1/URA3 rate divided by the yel068c::CAN1/URA3 rate.

rad51Δ rad59Δ double mutant, had a similar GCR rate in the yel072w::CAN1/URA3 assay and a lower rate in the yel068c::CAN1/URA3 assay. In addition, the rate of t(V;XIV) translocations was reduced in the pol32Δ rad59Δ mutant relative to pol32Δ and rad59Δ single mutants, but not to the extent seen in rad51Δ rad59Δ double mutants (Fig. 2b). These results suggest that POL32 functions in the RAD51-dependent pathway but not the RAD59-dependent pathway that promotes duplication-mediated GCRs; however, in the RAD51-dependent pathway, the formation of duplication-mediated GCRs is not completely dependent on POL32. Thus, a subset of the RAD51-dependent duplication-mediated GCRs is probably produced by POL32-dependent BIR, whereas POL32-independent RAD51-dependent and RAD59-dependent duplication-mediated GCRs either result from other HR mechanisms, such as a half-crossover mechanism[20], or are produced by a BIR pathway that has different genetic requirements than BIR driven by HO-induced DSBs. Two other replication-associated mutations, pri2-1, which suppresses HR-mediated BIR[32], and pol12-100, which increases levels of Holliday junctions during replication[33], generally decreased or weakly increased GCR rates, respectively.

Pathways that suppress HR-mediated GCRs

Because analysis of sgs1Δ, top3Δ, rmi1Δ, msh2Δ and msh6Δ mutants (Table 2) indicated that the yel072w::CAN1/URA3 assay can reveal pathways that specifically suppress duplication-mediated rearrangements, we screened for additional context-specific mutations (Table 3). Deletion of SGS1 causes synthetic growth defects with deletions of SLX1, SLX4, SLX5, SLX8, MUS81, SAE2, SRS2 or RRM3. Deletion of each of these genes, except SAE2, caused duplication-specific increases in GCR rates, whereas only deletion of SAE2 and MUS81 caused increases in GCRs mediated by single-copy DNA sequences. Similarly, deletion of the repair genes RAD6, MPH1, RAD10 or EXO1 caused large increases in duplication-specific GCR rates, but little or no increase in single-copy-sequence-mediated GCRs. The duplication-specific effects of rad10Δ contrast with previous findings that the Rad1–Rad10 complex is required for single-copy DNA-sequence-mediated GCRs[34]. Deletion of ESC2 and ESC4 (also called RTT107), which encodes a protein recruited to stalled replication forks[35], caused a general increase in GCR rate and a preferential increase in the rate of HXT13-DSF1 duplication-mediated GCRs. Defects in chromatin modifying pathways caused by deletion of ASF1, RTT109, ARP8 or NHP10 also had duplication-specific effects; however, in contrast to deleting ASF1 or RTT109, deleting ARP8 and NHP10, which encode subunits of the Ino80

chromatin remodelling complex[36], did not alter the rate of single-copy-sequence-mediated GCRs. In contrast, deletion of CTF18, which causes sister chromatid cohesion defects[37], caused similar increases in both assays. These results demonstrate that the genetics of suppressing GCRs changes substantially depending on chromosomal features in the breakpoint region.

Checkpoint suppression of HR-mediated GCRs

Deletion of MRC1, which encodes a Rad53 coactivator with roles in DNA replication and replication stress checkpoint signalling[38], caused a small increase in the rate of single-copy sequence-mediated GCRs and a large increase in HXT13-DSF1 duplication-mediated GCRs. The latter GCRs were primarily homology-driven translocations (Fig. 2a), and two GCRs predicted to be t(V;XIV) translocations by PCR were non-reciprocal translocations (Supplementary Fig. 3) similar to all other duplication-mediated GCRs analysed by aCGH (Fig. 1c and Supplementary Figs 2 and 3). The mrc1-aq allele, which specifically affects the checkpoint function of MRC1 (ref. 38), had little effect on single-copy-sequence-mediated GCRs but caused a large increase in duplication-mediated GCRs (Table 3). Similarly, deleting TOF1, which encodes another replication fork and checkpoint protein[39], caused a specific increase in HXT13-DSF1 duplication-mediated GCRs (Table 3). We found a synergistic interaction between mrc1Δ and tof1Δ but not between mrc1-aq and tof1Δ in both the yel068c::CAN1/URA3 and yel072w::CAN1/URA3 assays, indicating a partial redundancy of these genes (Table 3).

Mutations in the checkpoint genes RAD24, MEC1, RAD53, DUN1 and CHK1 increased the GCR rate in both the yel068c::CAN1/URA3 and yel072w::CAN1/URA3 assays (Table 3), although the affect on duplication-mediated GCR rates was possibly not as large as that of mrc1Δ or tof1Δ mutations, raising the possibility that mrc1Δ and tof1Δ mutations might increase DNA damage in addition to causing checkpoint defects. Mutations in TEL1, which encodes a protein kinase that is partially redundant with Mec1, resulted in small rate increases in both GCR assays, consistent with a small checkpoint role for Tel1 in the presence of Mec1; however, tel1Δ telomere maintenance defects could contribute to a low level of GCRs. Mutations in RAD9, which encodes an alternative Rad53 co-activator that responds to general DNA damage signalling, but not replication fork damage in strains with MRC1 (ref. 39), were similar to the affects of damage checkpoint mutations on single-copy-sequence-mediated GCRs, but caused a much smaller increase than these mutations in the rate of duplication-mediated GCRs. Together, these data suggest that the DNA damage checkpoint primarily suppresses single-copy-sequence-mediated

Table 3 | Effect of mutations on the accumulation of duplication-mediated rearrangements

Genotype	yel068c:: CAN1/URA3		yel072w:: CAN1/URA3		Ratio†
	Strain	GCR rate*	Strain	GCR rate*	
Wild type	RDKY6677	2.27×10^{-9} (5.1)	RDKY6678	1.97×10^{-8} (56)	8.7
SGS1 interactors					
mus81Δ	RDKY6731	1.26×10^{-8} (36)	RDKY6748	2.51×10^{-7} (717)	20
rrm3Δ	RDKY6735	9.46×10^{-10} (2.7)	RDKY6752	3.87×10^{-8} (110)	41
sae2Δ	RDKY6737	4.23×10^{-8} (120)	RDKY6754	1.65×10^{-7} (470)	3.9
slx1Δ	RDKY6738	$<1.12 \times 10^{-9}$ (<3.2)	RDKY6755	2.32×10^{-8} (66)	>20.6
slx4Δ	RDKY6739	$<7.94 \times 10^{-10}$ (<2.3)	RDKY6756	9.26×10^{-8} (264)	>116
slx5Δ	RDKY6740	1.48×10^{-9} (4.2)	RDKY6757	4.82×10^{-7} (1,378)	326
slx8Δ	RDKY6846	$<1.81 \times 10^{-9}$ (<5.2)	RDKY6847	9.65×10^{-7} (2,757)	>532
srs2Δ	RDKY6741	7.18×10^{-10} (2.1)	RDKY6758	1.28×10^{-7} (365)	178
Chromatin					
asf1Δ	RDKY6725	1.34×10^{-8} (38)	RDKY6742	2.89×10^{-7} (825)	22
arp8Δ	RDKY6726	$<6.05 \times 10^{-10}$ (<1.73)	RDKY6743	4.84×10^{-8} (138)	>80
nhp10Δ	RDKY6732	1.39×10^{-9} (4.0)	RDKY6749	3.01×10^{-8} (86)	22
rtt109Δ	RDKY6736	5.64×10^{-9} (16)	RDKY6753	1.84×10^{-7} (526)	33
Cohesion					
ctf18Δ	RDKY6727	2.40×10^{-8} (69)	RDKY6744	2.22×10^{-7} (633)	9.2
Other repair					
esc2Δ	RDKY6878	4.36×10^{-8} (124)	RDKY6879	1.07×10^{-5} (30,700)	247
esc4Δ	RDKY6728	1.66×10^{-8} (48)	RDKY6745	3.07×10^{-7} (876)	18.5
exo1Δ	RDKY6729	2.00×10^{-9} (5.7)	RDKY6746	8.44×10^{-8} (241)	42
mph1Δ	RDKY6794	2.00×10^{-9} (5.7)	RDKY6795	1.05×10^{-7} (300)	53
rad6Δ	RDKY6733	4.66×10^{-9} (13)	RDKY6750	6.03×10^{-7} (1,724)	130
rad10Δ	RDKY6734	8.49×10^{-10} (2.4)	RDKY6751	1.80×10^{-7} (404)	212
Checkpoint					
mrc1Δ	RDKY6730	3.35×10^{-9} (9.6)	RDKY6747	3.75×10^{-7} (1,071)	112
mrc1-aq	RDKY6766	1.51×10^{-9} (4.3)	RDKY6775	1.23×10^{-7} (351)	81
tof1Δ	RDKY6767	5.71×10^{-9} (16)	RDKY6776	4.25×10^{-7} (1,214)	74
mrc1Δ tof1Δ	RDKY6779	6.41×10^{-8} (183)	RDKY6780	1.26×10^{-6} (3,612)	20
mrc1-aq tof1Δ	RDKY6849	3.69×10^{-9} (11)	RDKY6849	2.06×10^{-7} (589)	56
rad24Δ	RDKY6759	2.00×10^{-8} (57.3)	RDKY6768	1.97×10^{-7} (555)	9.7
mec1Δ sml1Δ	RDKY6760	2.34×10^{-8} (67)	RDKY6769	1.50×10^{-7} (429)	6.4
tel1Δ	RDKY6761	4.99×10^{-9} (14)	RDKY6770	2.87×10^{-8} (82)	5.8
rad53Δ sml1Δ	RDKY6762	5.60×10^{-8} (160)	RDKY6771	3.05×10^{-7} (871)	5.4
rad9Δ	RDKY6765	2.17×10^{-8} (62)	RDKY6774	3.82×10^{-8} (109)	1.8
chk1Δ	RDKY6764	1.76×10^{-8} (50)	RDKY6773	1.96×10^{-7} (560)	11
dun1Δ	RDKY6763	1.63×10^{-8} (47)	RDKY6772	1.61×10^{-7} (461)	9.9

* Rate of accumulating Can^r 5-FOA^r progeny. The number in parentheses is the fold increase relative to the standard wild-type rate (3.5×10^{-10}; ref. 16).
† The *yel072w::CAN1/URA3* rate divided by the *yel068c::CAN1/URA3* rate.

GCRs, whereas both the DNA damage checkpoint, to a lesser extent, and the replication stress checkpoint, to a much greater extent, suppress duplication-mediated GCRs.

Discussion

We have found that many genes have little or no role in suppressing GCRs in single-copy sequences but have a large role in suppressing GCRs mediated by non-allelic HR at the *HXT13-DSF1* at-risk sequence that resembles a segmental duplication in mammalian cells. One group of genes includes the MMR genes and the genes encoding the Sgs1–Top3–Rmi1 complex that suppress HR between divergent sequences[29]. Another group included *MRC1* and *TOF1*, and our analysis of checkpoint genes indicated that the replication stress checkpoint is critical in suppressing *HXT13-DSF1*-mediated GCRs but not single-copy-sequence-mediated GCRs. A third group of genes that almost all exclusively function in suppressing *HXT13-DSF1* duplication-mediated GCRs include *SRS2*, *RRM3*, *MUS81*, *SAE2*, *SLX1*, *SLX4*, *SLX5* and *SLX8*, which cause synthetic growth defects when deleted in combination with an *sgs1Δ* mutation, owing to accumulation of toxic replication intermediates that in many cases can be suppressed by a HR defect[40–43]. Potentially related to these genes are: *RAD10* and *EXO1* which encode an endonuclease and an exonuclease, respectively, that can act in processing of HR and aberrant replication intermediates[44,45]; *MPH1* encoding a DNA helicase which may disrupt HR intermediates like Sgs1 (refs 46–48); and *RAD6* which regulates processes that act on replication forks that encounter DNA damage[49]. Finally, *ARP8*, *NHP10*, *ASF1* and *RTT109*, which function in chromatin remodelling and checkpoint regulation and can act during S phase[36,50], strongly suppressed duplication-mediated GCRs. All of these genes may function in

responses to replication stress, including checkpoint activation or shut off, repair of aberrant replication intermediates and suppression of the formation of aberrant replication intermediates, and some clearly act to directly prevent aberrant HR. How might the products of these genes act so specifically to prevent duplication-mediated GCRs? It is unlikely that they solely act to prevent aberrant DNA structures during replication such as DSBs as they would also suppress GCRs mediated by single-copy DNA sequences. Rather, they may prevent aberrant HR such as homeologous recombination or aberrant BIR intermediates so that HR can selectively target homologous sequences on sister chromatids and homologues as well as restart damaged replication forks to prevent genome instability rather than result in HR-mediated GCRs.

Our results indicate that dispersed repetitive elements in DNA resembling segmental duplications are at risk for causing genomic instability. The presence of multiple pathways that are highly specific for suppressing rearrangements between these elements explains how genomes remain stable despite the presence of sequences at risk for mediating genome rearrangements. These results complement previous studies that identified critical pathways and genes that suppress GCRs that target single-copy sequences[17]. Overall, our data suggest that defects in different DNA repair pathways result in distinct GCR signatures that may be diagnostic of the defects that underlie genome instability.

METHODS SUMMARY

Yeast strains were constructed by deleting *CAN1* and integrating a telomeric hygromycin marker and a *CAN1/URA3* cassette in the RDKY3023 background (*MATa leu2Δ1 his3Δ200 trp1Δ63 lys2ΔBgl hom3-10 ade2Δ1 ade8 ura3-52*). GCRs were selected using standard methods[16]. GCR products were analysed by PCR and by aCGH (NimbleGen).

1. Online Mendelian Inheritance in Man, OMIM. Institute of Genetic Medicine, Johns Hopkins University and National Center for Biotechnology Information, National Library of Medicine. ⟨http://www.ncbi.nlm.nih.gov/omim/⟩ (1999).
2. Stankiewicz, P. & Lupski, J. R. The genomic basis of disease, mechanisms and assays for genomic disorders. *Genome Dyn.* **1**, 1–16 (2006).
3. Mitelman, F. *Catalog of Chromosome Aberrations in Cancer* (Wiley Liss, 1991).
4. Gorringe, K. L. *et al.* Evidence that both genetic instability and selection contribute to the accumulation of chromosome alterations in cancer. *Carcinogenesis* **26**, 923–930 (2005).
5. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).
6. Ribas, M. *et al.* The structural nature of chromosomal instability in colon cancer cells. *FASEB J.* **17**, 289–291 (2003).
7. Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
8. Deininger, P. L. & Batzer, M. A. Alu repeats and human disease. *Mol. Genet. Metab.* **67**, 183–193 (1999).
9. Gordenin, D. A. & Resnick, M. A. Yeast ARMs (DNA at-risk motifs) can reveal sources of genome instability. *Mutat. Res.* **400**, 45–58 (1998).
10. Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nature Rev. Genet.* **3**, 370–379 (2002).
11. Ji, Y., Eichler, E. E., Schwartz, S. & Nicholls, R. D. Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.* **10**, 597–610 (2000).
12. Harris, S., Rudnicki, K. S. & Haber, J. E. Gene conversions and crossing over during homologous and homeologous ectopic recombination in *Saccharomyces cerevisiae*. *Genetics* **135**, 5–16 (1993).
13. Umezu, K., Hiraoka, M., Mori, M. & Maki, H. Structural analysis of aberrant chromosomes that occur spontaneously in diploid *Saccharomyces cerevisiae*: retrotransposon Ty1 plays a crucial role in chromosomal rearrangements. *Genetics* **160**, 97–110 (2002).
14. Lemoine, F. J., Degtyareva, N. P., Lobachev, K. & Petes, T. D. Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* **120**, 587–598 (2005).
15. Lobachev, K. S. *et al.* Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.* **19**, 3822–3830 (2000).
16. Chen, C. & Kolodner, R. D. Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nature Genet.* **23**, 81–85 (1999).
17. Putnam, C. D., Pennaneach, V. & Kolodner, R. D. *Saccharomyces cerevisiae* as a model system to define the chromosomal instability phenotype. *Mol. Cell. Biol.* **25**, 7226–7238 (2005).
18. Eichler, E. E. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.* **17**, 661–669 (2001).
19. Bosco, G. & Haber, J. E. Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* **150**, 1037–1047 (1998).
20. Deem, A. *et al.* Defective break-induced replication leads to half-crossovers in *Saccharomyces cerevisiae*. *Genetics* **179**, 1845–1860 (2008).
21. Boulton, S. J. & Jackson, S. P. Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J.* **17**, 1819–1828 (1998).
22. Smith, C. E., Llorente, B. & Symington, L. S. Template switching during break-induced replication. *Nature* **447**, 102–105 (2007).
23. Schmidt, K. H., Wu, J. & Kolodner, R. D. Control of translocations between highly diverged genes by Sgs1, the *Saccharomyces cerevisiae* homolog of the Bloom's syndrome protein. *Mol. Cell. Biol.* **26**, 5406–5420 (2006).
24. Oh, S. D. *et al.* BLM ortholog, Sgs1, prevents aberrant crossing-over by suppressing formation of multichromatid joint molecules. *Cell* **130**, 259–272 (2007).
25. Myung, K., Chen, C. & Kolodner, R. D. Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* **411**, 1073–1076 (2001).
26. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233–271 (2004).
27. Bai, Y. & Symington, L. S. A Rad52 homolog is required for RAD51-independent mitotic recombination in *Saccharomyces cerevisiae*. *Genes Dev.* **10**, 2025–2037 (1996).
28. Spell, R. M. & Jinks-Robertson, S. Examination of the roles of Sgs1 and Srs2 helicases in the enforcement of recombination fidelity in *Saccharomyces cerevisiae*. *Genetics* **168**, 1855–1865 (2004).
29. Myung, K., Datta, A., Chen, C. & Kolodner, R. D. SGS1, the *Saccharomyces cerevisiae* homologue of BLM and WRN, suppresses genome instability and homeologous recombination. *Nature Genet.* **27**, 113–116 (2001).
30. Sugawara, N., Goldfarb, T., Studamire, B., Alani, E. & Haber, J. E. Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. *Proc. Natl Acad. Sci. USA* **101**, 9315–9320 (2004).
31. Mullen, J. R., Nallaseth, F. S., Lan, Y. Q., Slagle, C. E. & Brill, S. J. Yeast Rmi1/Nce4 controls genome stability as a subunit of the Sgs1-Top3 complex. *Mol. Cell. Biol.* **25**, 4476–4487 (2005).
32. Lydeard, J. R., Jain, S., Yamaguchi, M. & Haber, J. E. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* **448**, 820–823 (2007).
33. Zou, H. & Rothstein, R. Holliday junctions accumulate in replication mutants via a RecA homolog-independent mechanism. *Cell* **90**, 87–96 (1997).
34. Hwang, J. Y., Smith, S. & Myung, K. The Rad1-Rad10 complex promotes the production of gross chromosomal rearrangements from spontaneous DNA damage in *Saccharomyces cerevisiae*. *Genetics* **169**, 1927–1937 (2005).
35. Roberts, T. M., Zaidi, I. W., Vaisica, J. A., Peter, M. & Brown, G. W. Regulation of Rtt107 recruitment to stalled DNA replication forks by the cullin Rtt101 and the Rtt109 acetyltransferase. *Mol. Biol. Cell* **19**, 171–180 (2008).
36. Conaway, R. C. & Conaway, J. W. The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem. Sci.* **34**, 71–77 (2009).
37. Mayer, M. L., Gygi, S. P., Aebersold, R. & Hieter, P. Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in *S. cerevisiae*. *Mol. Cell* **7**, 959–970 (2001).
38. Osborn, A. J. & Elledge, S. J. Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes Dev.* **17**, 1755–1767 (2003).
39. Katou, Y. *et al.* S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature* **424**, 1078–1083 (2003).
40. Mullen, J. R., Kaliraman, V., Ibrahim, S. S. & Brill, S. J. Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in *Saccharomyces cerevisiae*. *Genetics* **157**, 103–118 (2001).
41. Schmidt, K. H. & Kolodner, R. D. Requirement of Rrm3 helicase for repair of spontaneous DNA lesions in cells lacking Srs2 or Sgs1 helicase. *Mol. Cell. Biol.* **24**, 3213–3226 (2004).
42. Fabre, F., Chan, A., Heyer, W. D. & Gangloff, S. Alternate pathways involving Sgs1/Top3, Mus81/Mms4, and Srs2 prevent formation of toxic recombination intermediates from single-stranded gaps created by DNA replication. *Proc. Natl Acad. Sci. USA* **99**, 16887–16892 (2002).
43. Torres, J. Z., Schnakenberg, S. L. & Zakian, V. A. *Saccharomyces cerevisiae* Rrm3p DNA helicase promotes genome integrity by preventing replication fork stalling: viability of rrm3 cells requires the intra-S-phase checkpoint and fork restart activities. *Mol. Cell. Biol.* **24**, 3198–3212 (2004).
44. Sugawara, N., Paques, F., Colaiacovo, M. & Haber, J. E. Role of *Saccharomyces cerevisiae* Msh2 and Msh3 repair proteins in double-strand break-induced recombination. *Proc. Natl Acad. Sci. USA* **94**, 9214–9219 (1997).
45. Fiorentini, P., Huang, K. N., Tishkoff, D. X., Kolodner, R. D. & Symington, L. S. Exonuclease I of *Saccharomyces cerevisiae* functions in mitotic recombination *in vivo* and *in vitro*. *Mol. Cell. Biol.* **17**, 2764–2773 (1997).
46. Krejci, L. *et al.* DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* **423**, 305–309 (2003).
47. Prakash, R. *et al.* Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination. *Genes Dev.* **23**, 67–79 (2009).
48. Veaute, X. *et al.* The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309–312 (2003).
49. Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G. & Jentsch, S. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**, 135–141 (2002).
50. Chen, C. C. *et al.* Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. *Cell* **134**, 231–243 (2008).

Author Contributions C.D.P., R.D.K. and T.K.H. designed the experiments. C.D.P. and T.K.H. performed the experiment. C.D.P. and R.D.K. analysed the data and wrote the manuscript.

Author Information Microarray data have been submitted to ArrayExpress (http://www.ebi.ac.uk/arrayexpress) with accession number E-TABM-714. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.D.K. (rkolodner@ucsd.edu).

METHODS

Plasmid construction. A *can1::hisG-URA3-hisG* disruption cassette was constructed by first PCR-amplifying fragments that are telomeric to *CAN1* (chromosome V 30187–30928) flanked by ApaI and XhoI sites and centromeric to *CAN1* (chromosome V 34339–34965) flanked by XbaI and BamHI sites and inserting them into pRS315 (ref. 51) to generate pRDK1374. A *hisG-URA3-hisG* fragment was amplified from pNKY51 (ref. 52) and was then inserted into SmaI-digested pRS315 by recombinational cloning in *S. cerevisiae*. Then the *hisG-URA3-hisG* fragment was subcloned into pRDK1374 between SalI and BamHI sites to generate pRDK1375 containing the *hisG-URA3-hisG* fragment flanked by 626 bp of upstream and 741 bp of downstream homology to the *CAN1* locus.

The *CAN1/URA3* cassette was constructed by cloning fragments of *CAN1* and *URA3* into a plasmid with flanking NheI sites. The *CAN1* gene and flanking sequence (chromosome V 30952–34315) was amplified by PCR and cloned into pCR2.1-TOPO (Invitrogen) to generate pRDK1376. The *URA3* gene and flanking sequence (chromosome V 116011–117061) was amplified by PCR with primers to introduce flanking XbaI sites, and cloned into pRDK1376; inserts with *CAN1* and *URA3* in divergent orientations were selected. The *CAN1/URA3* cassette was then PCR amplified with primers adding flanking NheI sites, cloned into pCRT7CT (Invitrogen), and verified by sequencing to generate pRDK1377.

For each chromosome V integration site, integration constructs were generated by subcloning the pRDK1377 NheI fragment into plasmids containing the target genes of interest. The gene and flanking regions of *YEL072W* (chromosome V 12961–14898) and *YEL068C* (chromosome V 25222–26411) were amplified by PCR, cloned into pRS315 and modified by site-directed mutagenesis to introduce NheI sites into the centre of the genes. Subcloning the *CAN1/URA3* cassette into the engineered NheI sites in *YEL072W* and *YEL068C* generated the plasmids pRDK1378 and pRDK1379, respectively. Similarly, the gene and flanking regions of *YEL064C* (chromosome V 30060–30928) and *YEL062W* (chromosome V 36007–36992) were amplified by PCR, cloned into pET21a and the NheI-digested *CAN1/URA3* cassette was subcloned into compatible SpeI sites to generate plasmids pRDK1380 and pRDK1381, respectively.

Genetic methods. YPD and synthetic drop-out media for propagation of strains have been described previously[16]. The *can1::hisG-URA3-hisG* integration fragment was cut out from pRDK1375 using KpnI and SacI and transformed into RDKY3023 (*MATa leu2Δ1 his3Δ200 trp1Δ63 lys2ΔBgl hom3-10 ade2Δ1 ade8 ura3-52*). Uracil prototrophs were verified by PCR, and a *can1::hisG* uracil auxotroph, RDKY5461, was selected on 5-FOA-containing medium. The *CAN1/URA3* integration cassettes were amplified by PCR from plasmids described above and integrated into RDKY5461. These strains were then modified by integrating a hygromycin-resistance cassette telomeric to *YEL072W* (chromosome V 11081–11618) to generate RDKY6678 (*yel072w::CAN1/URA3*), RDKY6677 (*yel068c::CAN1/URA3*), RDKY6676 (*yel064c::CAN1/URA3*) and RDKY6675 (*yel062w::CAN1/URA3*). Additional mutations were added to these strains using standard PCR-based mutagenesis methods, pop-in, pop-out plasmid vectors or intercrossing with mutants derived from the same parental strain background (Supplementary Table 2). The media and protocol for measuring GCR rates were essentially as described previously[16]. Ninety-five per cent confidence intervals of the median were calculated by a two-sided nonparametric test (http://www.math.unb.ca/~knight/utility/MedInt95.htm). The GCR rates determined using this method are highly reliable. Using data from a number of studies covering a broad range of mutants and GCR rates, we have calculated that the average upper and lower 95% confidence interval limits are 1.5 and 0.7, respectively, times the median GCR rate determined. In less than 8% of the measurements were the upper and lower 95% confidence intervals greater than 2 or less than 0.5, respectively, times the median GCR rate determined.

Analysis of GCR Isolates. GCR isolates were tested for loss of hygromycin resistance by growth on YPD media containing $300\,\mu g\,ml^{-1}$ hygromycin B (Invitrogen). Genomic DNA was prepared from individual isolates and subjected to PCR analysis to categorize GCRs. The t(V;XIV) and t(V;IV or X) translocations were identified by amplification of the junction region with a chromosome-V-specific primer centromeric to the *HXT13-DSF1* region and a chromosome XIV- or IV/X-specific primer telomeric to the *HXT13-DSF1* homologies on those chromosomes under conditions where no product was generated with DNA from wild-type strains. A series of PCR reactions spanning the ~20-kb region between *HXT13* and *PCM1* on chromosome V were used to map breakpoints for isolates that were not t(V;XIV) or t(V;IV or X) translocations by identifying the region where all telomeric reactions failed and all centromeric reactions succeeded. Breakpoint junctions from selected t(V;XIV) isolates were amplified as described above and sequenced by dye terminator DNA sequencing.

Array comparative genomic hybridization. A total of 1 μg of genomic DNA was prepared from GCR isolates and wild-type RDKY6678 using the Purgene kit (Qiagen) and concentrated to over $100\,ng\,\mu l^{-1}$. GCR isolate DNA was amplified and labelled with Cy5 and the wild-type DNA was amplified and labelled with Cy3 and GCR isolate/wild-type pairs were applied to a NimbleGen 4-plex chip. Data were analysed using the SignalMap software (NimbleGen).

51. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
52. Alani, E., Cao, L. & Kleckner, N. A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. *Genetics* **116**, 541–545 (1987).

LETTERS

An upper limit on the stochastic gravitational-wave background of cosmological origin

The LIGO Scientific Collaboration* & The Virgo Collaboration*

A stochastic background of gravitational waves is expected to arise from a superposition of a large number of unresolved gravitational-wave sources of astrophysical and cosmological origin. It should carry unique signatures from the earliest epochs in the evolution of the Universe, inaccessible to standard astrophysical observations[1]. Direct measurements of the amplitude of this background are therefore of fundamental importance for understanding the evolution of the Universe when it was younger than one minute. Here we report limits on the amplitude of the stochastic gravitational-wave background using the data from a two-year science run of the Laser Interferometer Gravitational-wave Observatory[2] (LIGO). Our result constrains the energy density of the stochastic gravitational-wave background normalized by the critical energy density of the Universe, in the frequency band around 100 Hz, to be $<6.9 \times 10^{-6}$ at 95% confidence. The data rule out models of early Universe evolution with relatively large equation-of-state parameter[3], as well as cosmic (super)string models with relatively small string tension[4] that are favoured in some string theory models[5]. This search for the stochastic background improves on the indirect limits from Big Bang nucleosynthesis[1,6] and cosmic microwave background[7] at 100 Hz.

According to the general theory of relativity, gravitational waves are produced by accelerating mass distributions with a quadrupole (or higher) moment. Moreover, in the early phases of the evolution of the Universe, they can be produced by the mechanism of amplification of vacuum fluctuations. Once produced, gravitational waves travel through space-time at the speed of light, and are essentially unaffected by the matter they encounter. As a result, gravitational waves emitted shortly after the Big Bang (and observed today) would carry unaltered information about the physical processes that generated them. These waves are expected to be generated by a large number of unresolved sources, forming a stochastic gravitational-wave background (SGWB) that is usually described in terms of the gravitational-wave spectrum:

$$\Omega_{GW}(f) = \frac{f}{\rho_c} \frac{d\rho_{GW}}{df} \qquad (1)$$

where $d\rho_{GW}$ is the energy density of gravitational radiation contained in the frequency range f to $f + df$ and ρ_c is the critical energy density of the Universe[8]. Many cosmological mechanisms for generation of the SGWB exist, such as the inflationary models[9,10], pre-Big-Bang models[11–13], electroweak phase transition[14], and cosmic strings[4,5,15,16]. There are also astrophysical mechanisms, such as magnetars[17] or rotating neutron stars[18].

The physical manifestation of gravitational waves consists of stretching and compressing the spatial dimensions orthogonal to the direction of wave propagation, producing strain in an oscillating quadrupolar pattern. A Michelson interferometer with suspended

mirrors[2] is well suited to measure this differential strain signal due to gravitational waves. Over the past decade, LIGO has built three such multi-kilometre interferometers, at two locations[2]: H1 (4 km) and H2 (2 km) share the same facility at Hanford, Washington, USA, and L1 (4 km) is located in Livingston Parish, Louisiana, USA. LIGO, together with the 3 km interferometer Virgo[19] in Italy and GEO[20] in Germany, forms a network of gravitational-wave observatories. LIGO has completed science run S5 (between 5 November 2005 and 30 September 2007), acquiring one year of data coincident among H1, H2 and L1, at the interferometer design sensitivities (Fig. 1).

The search for the SGWB using LIGO data is performed by cross-correlating strain data from pairs of interferometers[8]. In the frequency (f) domain, the cross-correlation between two interferometers is multiplied by a filter function $\tilde{Q}(f)$ (Supplementary Information):

$$\tilde{Q}(f) = N \frac{\gamma(f) \Omega_{GW}(f) H_0^2}{f^3 P_1(f) P_2(f)} \qquad (2)$$
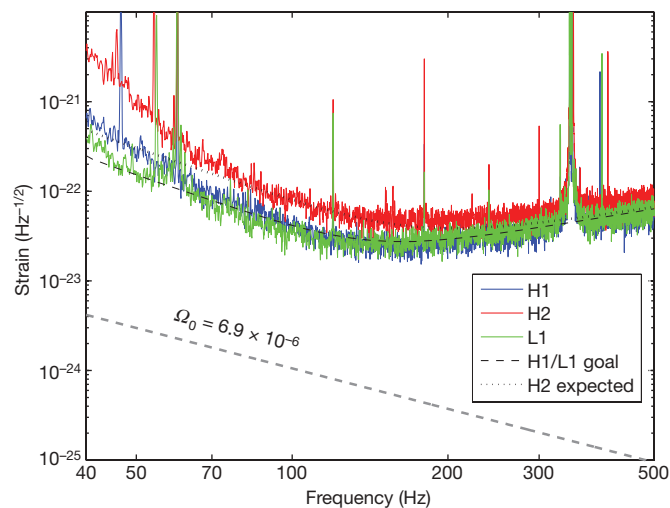
Figure 1 | Sensitivities of LIGO interferometers. LIGO interferometers reached their design sensitivity in November 2005, resulting in interferometer strain noise at the level of 3×10^{-22} r.m.s. in a 100 Hz band around 100 Hz. This figure shows typical strain sensitivities of LIGO interferometers during the subsequent science run S5. Also shown is the strain amplitude corresponding to the upper limit on the gravitational-wave energy density presented in this paper (grey dashed line). Note that this upper limit is ~100 times lower than the individual interferometer sensitivities, which illustrates the advantage of using the cross-correlation technique in this analysis.

*Lists of participants and their affiliations appear at the end of the paper.

This filter optimizes the signal-to-noise ratio, enhancing the frequencies at which the signal of the template gravitational-wave spectrum $\Omega_{GW}(f)$ is strong, while suppressing the frequencies at which the detector noise ($P_1(f)$ and $P_2(f)$) is large. In equation (2), and throughout this Letter, we assume the present value of the Hubble parameter $H_0 = 72$ km s^{-1} Mpc^{-1} (ref. 21), and use $\gamma(f)$ to denote the overlap reduction function[8], arising from the overlap of antenna patterns of interferometers at different locations and with different orientations. For the H1–L1 and H2–L1 pairs, the sensitivity above roughly 50 Hz is attenuated due to the overlap reduction. As most theoretical models in the LIGO frequency band are characterized by a power-law spectrum, we assume a power-law template gravitational-wave spectrum with index α: $\Omega_{GW}(f) = \Omega_\alpha (f/100\,\text{Hz})^\alpha$. The normalization constant N in equation (2) is chosen such that the expected value of the optimally filtered cross-correlation is Ω_α.

We apply the above search technique to the data acquired by LIGO during the science run S5. We include two interferometer pairs: H1–L1 and H2–L1. Summing up the contributions to the cross-correlation in the frequency band 41.5–169.25 Hz, which contains 99% of the sensitivity, leads to the final point estimate for the frequency independent gravitational-wave spectrum ($\alpha = 0$): $\Omega_0 = (2.1 \pm 2.7) \times 10^{-6}$, where the quoted error is statistical. We calculate the Bayesian 95% confidence upper limit for Ω_0, using the previous LIGO result (S4 run[22]) as a prior for Ω_0 and averaging over the interferometer calibration uncertainty. This procedure yields the 95% confidence upper limit $\Omega_0 < 6.9 \times 10^{-6}$. For other values of the power index α in the range between -3 and 3, the 95% upper limit varies between 1.9×10^{-6} and 7.1×10^{-6}. These results constitute more than an

order of magnitude improvement over the previous LIGO result in this frequency region[22]. Figure 2 shows this result in comparison with other observational constraints and some of the cosmological SGWB models.

Before the result described here, the most constraining bounds on the SGWB in the frequency band around 100 Hz came from the Big Bang nucleosynthesis (BBN) and from cosmic microwave background (CMB) measurements. The BBN bound is derived from the fact that a large gravitational-wave energy density at the time of BBN would alter the abundances of the light nuclei produced in the process. Hence, the BBN model and observations constrain the total gravitational-wave energy density at the time of nucleosynthesis[1,6]:

$$\Omega_{BBN} = \int \Omega_{GW}(f)\, d(\ln f) < 1.1 \times 10^{-5}(N_\nu - 3) \qquad (3)$$

where N_ν (the effective number of neutrino species at the time of BBN) captures the uncertainty in the radiation content during BBN. Measurements of the light-element abundances, combined with the Wilkinson Microwave Anisotropy Probe (WMAP) data give the upper bound $N_\nu - 3 < 1.4$ (ref. 23). Similarly, a large gravitational-wave background at the time of decoupling of CMB would alter the observed CMB and matter power spectra. Assuming homogeneous initial conditions, the total gravitational-wave energy density at the time of CMB decoupling is constrained to $\int \Omega_{GW}(f)\, d(\ln f) < 1.3 \times 10^{-5}$ (ref. 7). In the LIGO frequency band and for $\alpha = 0$, these bounds become: $\Omega_0^{BBN} < 1.1 \times 10^{-5}$ and $\Omega_0^{CMB} < 9.5 \times 10^{-6}$. Our result has now surpassed these bounds,
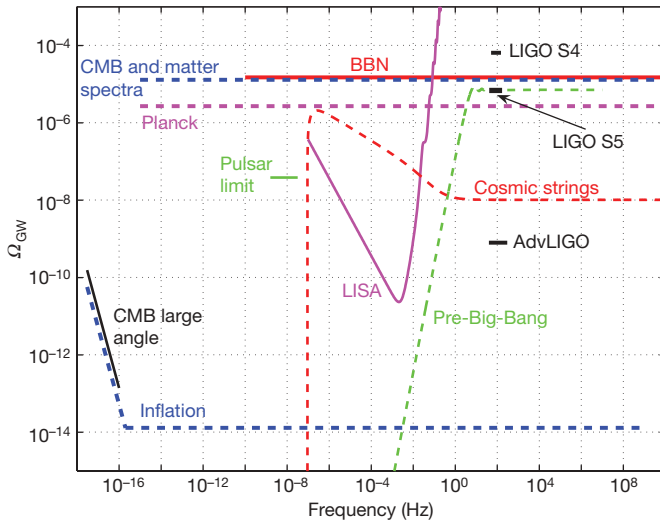
Figure 2 | Comparison of different SGWB measurements and models. The 95% upper limit presented here, $\Omega_0 < 6.9 \times 10^{-6}$ (LIGO S5), applies in the frequency band 41.5–169.25 Hz, and is compared to the previous LIGO S4 result[22] and to the projected Advanced LIGO sensitivity[25]. Note that the corresponding S5 95% upper bound on the total gravitational-wave energy density in this band, assuming frequency independent spectrum, is 9.7×10^{-6}. The indirect bound due to BBN[1,6] applies to $\Omega_{BBN} = \int \Omega_{GW}(f) d(\ln f)$ (and not to the density $\Omega_{GW}(f)$) over the frequency band denoted by the corresponding horizontal line, as defined in equation 3. A similar integral bound (over the range 10^{-15}–10^{10} Hz) can be placed using CMB and matter power spectra[7]. Projected sensitivities of the satellite-based Planck CMB experiment[7] and LISA gravitational-wave detector[26] are also shown. The pulsar bound[27] is based on the fluctuations in the pulse arrival times of millisecond pulsars and applies at frequencies around 10^{-8} Hz. Measurements of the CMB at large angular scales constrain the possible redshift of CMB photons due to the SGWB, and therefore limit the amplitude of the SGWB at largest wavelengths (smallest frequencies)[6]. Examples of inflationary[9,10], cosmic strings[4,5,15,16], and pre-Big-Bang[11–13] models are also shown (the amplitude and the spectral shape in these models can vary significantly as a function of model parameters).
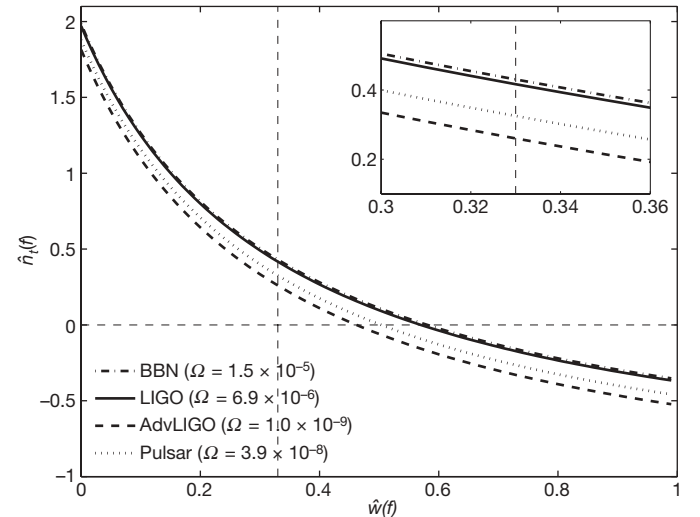
Figure 3 | Constraining early Universe evolution. The gravitational-wave spectrum $\Omega_{GW}(f)$ is related to the parameters that govern the evolution of the Universe[3]: $\Omega_{GW}(f) = A f^{\hat{a}(f)} f^{\hat{n}_t(f)} r$, where $\hat{\alpha}(f) = 2\dfrac{3\hat{w}(f) - 1}{3\hat{w}(f) + 1}$, r is the ratio of tensor and scalar perturbation amplitudes (measured by the CMB experiments), $\hat{n}_t(f)$ and $\hat{w}(f)$ are effective (average) tensor tilt and equation of state parameters respectively, and A is a constant depending on various cosmological parameters. Hence, the measurements of Ω_{GW} and r can be used to place constraints in the $\hat{w} - \hat{n}_t$ plane, independently of the cosmological model. The figure shows the $\hat{w} - \hat{n}_t$ plane for $r = 0.1$. The regions excluded by the BBN[23], LIGO and pulsar[27] bounds are above the corresponding curves (the inset shows a zoom-in on the central part of the figure). The BBN curve was calculated in ref. 3. We note that the CMB bound[7] almost exactly overlaps with the BBN bound. Also shown is the expected reach of Advanced LIGO[25]. Note that these bounds apply to different frequency bands, so their direct comparison is meaningful only if $\hat{n}_t(f)$ and $\hat{w}(f)$ are frequency independent. We note that for the simplest single-field inflationary model that still agrees with the cosmological data, with potential $V(\phi) = m^2\phi^2/2$ (where ϕ is a scalar field of mass m), $r = 0.14$ and $n_t(100\,\text{Hz}) = -0.035$ (ref. 28), implying a LIGO bound on the equation-of-state parameter of $\hat{w}(100\,\text{Hz}) < 0.59$.

which is one of the major milestones that LIGO was designed to achieve. Moreover, the BBN and CMB bounds apply only to backgrounds generated before the BBN and the CMB decoupling, respectively, while the LIGO bound also probes the SGWB produced later (this is the case, for example, in models involving cosmic strings).

Our result also constrains models of early Universe evolution. Although the evolution of the Universe following the BBN is well understood, there is little observational data probing the evolution before BBN, when the Universe was less than one minute old. The gravitational-wave spectrum $\Omega_{GW}(f)$ carries information about exactly this epoch in the evolution. In particular, measuring $\Omega_{GW}(f)$ is the best way to test for the existence of currently unknown 'stiff' energy components in the early Universe[3], for which a small density variation is associated with a large pressure change, which could carry information about the physics of the inflationary era[24]. Figure 3 demonstrates how the result presented here can be used to constrain the existence of these new energy components.

Our result also constrains models of cosmic (super)strings. Cosmic strings were originally proposed as topological defects formed during phase transitions in the early Universe[15]. More recently, it was realized that fundamental strings may also be expanded to cosmological scales[5]. Hence, searching for cosmic strings may provide a unique and powerful window into string theory and into particle physics at the highest energy scales. Figure 4 shows that our result, along with other observations, can be used to constrain the parameters in the cosmic string models. Whereas our result
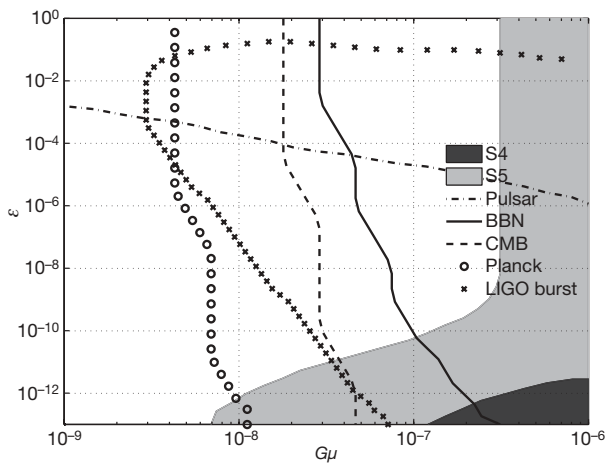
Figure 4 | Models involving cosmic strings. The network of cosmic strings is usually parametrized by the string tension μ (multiplied by the Newton constant G), and reconnection probability p. The CMB observations limit $G\mu < 10^{-6}$. If the size of the cosmic string loops is determined by the gravitational back-reaction[29], the size of the loop can be parametrized by a parameter ε (ref. 16), which is essentially unconstrained. The mechanism for production of gravitational waves relies on cosmic string cusps: regions of string that move at speeds close to the speed of light. If the cusp motion points towards Earth, a detectable burst of gravitational radiation may be produced[16,30]. The superposition of gravitational waves from all string cusps in the cosmic string network would produce a SGWB[4]. This figure shows how different experiments probe the $\varepsilon-G\mu$ plane for a typical value of $p = 10^{-3}$ (ref. 4) (p is expected to be in the range 10^{-4}–1). The excluded regions (always to the right of the corresponding curves) correspond to the S4 LIGO result[22], this result, the BBN bound[6,23], the CMB bound[7], and the pulsar limit[27]. In particular, the bound presented in this paper excludes a new region in this plane ($7 \times 10^{-9} < G\mu < 1.5 \times 10^{-7}$ and $\varepsilon < 8 \times 10^{-11}$), which is not accessible to any of the other measurements. Also shown is the expected sensitivity for the search for individual bursts from cosmic string cusps with LIGO S5 data[30]. The region to the right of this curve is expected to produce at least one cosmic string burst event detectable by LIGO during the S5 run. Note that this search is complementary to the search for the SGWB as it probes a different part of the parameter space. Also shown is the region that will be probed by the Planck satellite measurements of the CMB[7]. The entire plane shown here will be accessible to Advanced LIGO[25] SGWB search.
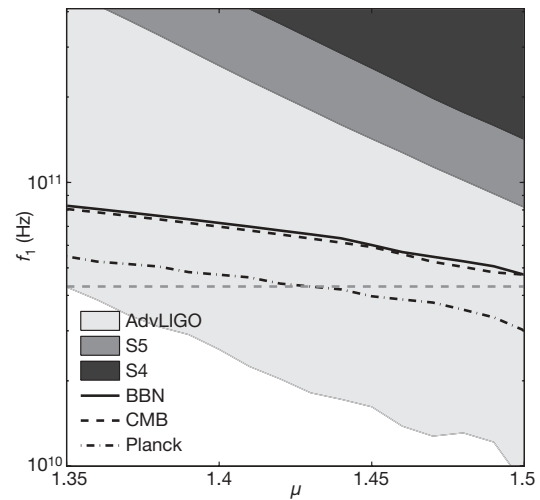
Figure 5 | Pre-Big-Bang models. In the pre-Big-Bang model, the gravitational waves are produced through the mechanism of amplification of vacuum fluctuations, analogously to the standard inflationary model. The typical gravitational-wave spectrum increases as f^3 up to a turn-over frequency f_s, above which $\Omega_{GW}(f) \propto f^{3-2\mu}$ with $\mu < 1.5$. The spectrum cuts off at a frequency f_1, which is theoretically expected to be within a factor of 10 from 4.3×10^{10} Hz (dashed horizontal line). This figure shows the $f_1-\mu$ plane for a representative value of $f_s = 30$ Hz. Excluded regions corresponding to the S4 result and to the result presented here are shaded. The regions excluded by the BBN[6,23] and the CMB[7] bounds are above the corresponding curves. The expected reaches of the Advanced LIGO[25] and of the Planck satellite[7] are also shown.

is currently excluding a fraction of the allowed parameter space, Advanced LIGO[25] is expected to probe most of these models.

Measurements of the SGWB also offer the possibility of probing alternative models of early Universe cosmology. For example, in the pre-Big-Bang model[11-13] the Universe starts off large and then undergoes a period of inflation driven by the kinetic energy of a dilaton field, after which the standard cosmology follows. Although more speculative than the standard cosmology model, the pre-Big-Bang model makes testable predictions of the gravitational-wave spectrum. As shown in Fig. 5, the BBN and CMB bounds are currently the most constraining for this model and Advanced LIGO[25] is expected to surpass them.

1. Maggiore, M. Gravitational wave experiments and early universe cosmology. *Phys. Rep.* **331**, 283–367 (2000).
2. Abbott, B. *et al.* Detector description and performance for the first coincidence observations between LIGO and GEO. *Nucl. Instrum. Meth. A* **517**, 154–179 (2004).
3. Boyle, L. & Buonanno, A. Relating gravitational wave constraints from primordial nucleosynthesis, pulsar timing, laser interferometers, and the CMB: implications for the early universe. *Phys. Rev. D* **78**, 043531 (2008).
4. Siemens, X., Mandic, V. & Creighton, J. Gravitational-wave stochastic background from cosmic strings. *Phys. Rev. Lett.* **98**, 111101 (2007).
5. Sarangi, S. & Tye, S. H. H. Cosmic string production towards the end of brane inflation. *Phys. Lett. B* **536**, 185–192 (2002).
6. Allen, B. The stochastic gravity-wave background: sources and detection. Preprint at <http://arXiv.org/abs/grqc/9604033> (1996).
7. Smith, T. L., Pierpaoli, E. & Kamionkowski, M. A new cosmic microwave background constraint to primordial gravitational waves. *Phys. Rev. Lett.* **97**, 021301 (2006).
8. Allen, B. & Romano, J. Detecting a stochastic background of gravitational radiation: signal processing strategies and sensitivities. *Phys. Rev. D* **59**, 102001 (1999).
9. Starobinskii, A. A. Spectrum of relict gravitational radiation and the early state of the universe. *JETP Lett.* **30**, 682–685 (1979).
10. Bar-Kana, R. Limits on direct detection of gravitational waves. *Phys. Rev. D* **50**, 1157–1162 (1994).
11. Brustein, R. *et al.* Relic gravitational waves from string cosmology. *Phys. Lett. B* **361**, 45–51 (1995).

12. Buonanno, A. et al. Spectrum of relic gravitational waves in string cosmology. Phys. Rev. D 55, 3330–3336 (1997).

13. Mandic, V. & Buonanno, A. Accessibility of the pre-big-bang models to LIGO. Phys. Rev. D 73, 063008 (2006).

14. Apreda, R. et al. Gravitational waves from electroweak phase transitions. Nucl. Phys. B 631, 342–368 (2002).

15. Kibble, T. W. B. Topology of cosmic domains and strings. J. Phys. A 9, 1387–1398 (1976).

16. Damour, T. & Vilenkin, A. Gravitational radiation from cosmic (super)strings: bursts, stochastic background, and observational windows. Phys. Rev. D 71, 063510 (2005).

17. Regimbau, T. & de Freitas Pacheco, J. A. Gravitational wave background from magnetars. Astron. Astrophys. 447, 1–8 (2006).

18. Regimbau, T. & de Freitas Pacheco, J. A. Cosmic background of gravitational waves from rotating neutron stars. Astron. Astrophys. 376, 381–385 (2001).

19. Acernese, F. et al. Status of Virgo. Class. Quant. Grav. 25, 114045 (2008).

20. Willke, B. et al. The GEO-HF project. Class. Quant. Grav. 23, S207–S214 (2006).

21. Bennet, C. L. et al. First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: preliminary maps and basic results. Astrophys. J. 148 (Suppl.), 1–28 (2003).

22. Abbott, B. et al. Searching for a stochastic background of gravitational waves with the Laser Interferometer Gravitational-Wave Observatory. Astrophys. J. 659, 918–930 (2007).

23. Cyburt, R. H. et al. New BBN limits on physics beyond the standard model from ^{4}He. Astropart. Phys. 23, 313–323 (2005).

24. Grishchuk, L. P. & Sidorov, Yu. V. Squeezed quantum states of relic gravitons and primordial density fluctuations. Phys. Rev. D 42, 3413–3421 (1990).

25. Advanced LIGO Team. Advanced LIGO reference design. LIGO preprint at <http://www.ligo.caltech.edu/docs/M/M060056-10.pdf> (2007).

26. Bender, P. L., Danzmann, K. & the LISA study team. Laser Interferometer Space Antenna for the Detection and Observation of Gravitational Waves: Pre-Phase A Report 2nd edn (MPQ233, Max-Plank Institut für Quantenoptik, 1998).

27. Jenet, F. A. et al. Upper bounds on the low-frequency stochastic gravitational wave background from pulsar timing observations: current limits and future prospects. Astrophys. J. 653, 1571–1576 (2006).

28. Komatsu, E. et al. Five-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation. Astrophys. J. 180 (Suppl.), 330–376 (2009).

29. Siemens, X. et al. Size of the smallest scales in cosmic string networks. Phys. Rev. D 66, 043501 (2002).

30. Siemens, X. et al. Gravitational wave bursts from cosmic (super)strings: quantitative analysis and constraints. Phys. Rev. D 73, 105001 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions are listed in Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to V.M. (mandic@physics.umn.edu).

The Ligo Scientific Collaboration and The VIRGO Collaboration

B. P. Abbott[1], R. Abbott[1], F. Acernese[2]‡||, R. Adhikari[1], P. Ajith[3], B. Allen[3,4], G. Allen[5], M. Alshourbagy[6]§, R. S. Amin[7], S. B. Anderson[1], W. G. Anderson[4], F. Antonucci[8]‡, S. Aoudia[9], M. A. Arain[10], M. Araya[1], H. Armandula[1], P. Armor[4], K. G. Arun[11], Y. Aso[1], S. Aston[12], P. Astone[8]‡, P. Aufmuth[13], C. Aulbert[3], S. Babak[14], P. Baker[15], G. Ballardin[16], S. Ballmer[1], C. Barker[17], D. Barker[17], F. Barone[2]‡||, B. Barr[18], P. Barriga[19], L. Barsotti[20], M. Barsuglia[21], M. A. Barton[1], I. Bartos[22], R. Bassiri[18], M. Bastarrika[18], Th. S. Bauer[23]‡, B. Behnke[14], M. Beker[23]§, M. Benacquista[24], J. Betzwieser[1], P. T. Beyersdorf[25], S. Bigotta[6]§, I. A. Bilenko[26], G. Billingsley[1], S. Birindelli[9], R. Biswas[4], M. A. Bizouard[11], E. Black[1], J. K. Blackburn[1], L. Blackburn[20], D. Blair[19], B. Bland[17], C. Boccara[27], T. P. Bodiya[20], L. Bogue[28], F. Bondu[9], L. Bonelli[6]§, R. Bork[1], V. Boschi[1], S. Bose[29], L. Bosi[30]‡, S. Braccini[6]‡, C. Bradaschia[6]‡, P. R. Brady[4], V. B. Braginsky[26], J. F. J. van den

Brand[23]‡§, J. E. Brau[31], D. O. Bridges[28], A. Brillet[9], M. Brinkmann[3], V. Brisson[11], C. Van Den Broeck[32], A. F. Brooks[1], D. A. Brown[33], A. Brummit[34], G. Brunet[20], A. Bullington[5], H.J. Bulten[23]‡§, A. Buonanno[35], O. Burmeister[3], R. L. Byer[5], L. Cadonati[37], G. Cagnoli[38]‡, E. Calloni[2]§, J. B. Camp[39], E. Campagna[38]‡||, J. Cannizzo[39], K. C. Cannon[1], B. Canuel[16], J. Cao[20], F. Carbognani[16], L. Cardenas[1], S. Caride[40], G. Castaldi[41], S. Caudill[7], M. Cavaglià[42], F. Cavalier[11], R. Cavalieri[16], G. Cella[6]‡, C. Cepeda[1], E. Cesarini[38]‡||, T. Chalermsongsak[1], E. Chalkley[18], P. Charlton[43], E. Chassande-Mottin[21], S. Chatterji[1,8]‡, S. Chelkowski[12], Y. Chen[14,44], N. Christensen[45], C. T. Y. Chung[46], D. Clark[5], J. Clark[32], J. H. Clayton[4], F. Cleva[9], E. Coccia[47]‡§, T. Cokelaer[32], C. N. Colacino[48], J. Colas[16], A. Colla[8]§, M. Colombini[8]§, R. Conte[49], D. Cook[17], T. R. C. Corbitt[20], C. Corda[6]§, N. Cornish[15], A. Corsi[8]‡, J.-P. Coulon[9], D. Coward[19], D. C. Coyne[1], J. D. E. Creighton[4], T. D. Creighton[24], A. M. Cruise[12], R. M. Culter[12], A. Cumming[18], L. Cunningham[18], E. Cuoco[16], S. L. Danilishin[26], S. D'Antonio[47]‡, K. Danzmann[3,13], A. Dari[30]‡§, V. Dattilo[16], B. Daudert[1], M. Davier[11], G. Davies[32], E. J. Daw[50], R. Day[16], R. De Rosa[2]§, D. DeBra[5], J. Degallaix[3], M. del Prete[6]‡||, V. Dergachev[40], S. Desai[51], R. DeSalvo[1], S. Dhurandhar[52], L. Di Fiore[2]‡, A. Di Lieto[6]§, M. Di Paolo Emilio[47]‡¶, A. Di Virgilio[6]‡, M. Díaz[24], A. Dietz[32], F. Donovan[20], K. L. Dooley[10], E. E. Doomes[53], M. Drago[54]||¶, R. W. P. Drever[55], J. Dueck[3], I. Duke[20], J.-C. Dumas[19], J. G. Dwyer[22], C. Echols[1], M. Edgar[18], A. Effler[17], P. Ehrens[1], G. Ely[45], E. Espinoza[1], T. Etzel[1], M. Evans[20], T. Evans[28], V. Fafone[47]§, S. Fairhurst[32], Y. Faltas[10], Y. Fan[19], D. Fazi[1], H. Fehrmann[3], I. Ferrante[6]§, F. Fidecaro[6]§, L. S. Finn[51], I. Fiori[16], R. Flaminio[56], K. Flasch[4], S. Foley[20], C. Forrest[57], N. Fotopoulos[4], J.-D. Fournier[9], J. Franc[56], A. Franzen[13], S. Frasca[8]‡§, F. Frasconi[6]‡, M. Frede[3], M. Frei[58], Z. Frei[48], A. Freise[12], R. Frey[31], T. Fricke[28], P. Fritschel[20], V. V. Frolov[28], M. Fyffe[28], V. Galdi[41], L. Gammaitoni[30]‡§, J. A. Garofoli[33], F. Garufi[2]§, E. Genin[16], A. Gennai[6]‡, I. Gholami[14], J. A. Giaime[7,28], S. Giampanis[3], K. D. Giardina[28], A. Giazotto[6]‡, K. Goda[20], E. Goetz[40], L. M. Goggin[4], G. González[7], M. L. Gorodetsky[26], S. Goßler[3], R. Gouaty[7], M. Granata[21], V. Granata[36], A. Grant[18], S. Gras[19], C. Gray[17], M. Gray[59], R. J. S. Greenhalgh[34], A. M. Gretarsson[60], C. Greverie[9], F. Grimaldi[20], R. Grosso[24], H. Grote[3], S. Grunewald[14], M. Guenther[17], G. Guidi[38]‡||, E. K. Gustafson[1], R. Gustafson[40], B. Hage[13], J. M. Hallam[12], D. Hammer[4], G. D. Hammond[18], C. Hanna[1], J. Hanson[28], J. Harms[61], G. M. Harry[20], I. W. Harry[32], E. D. Harstad[31], K. Haughian[18], K. Hayama[24], J. Heefner[1], H. Heitmann[9], P. Hello[11], I. S. Heng[18], A. Heptonstall[1], M. Hewitson[3], S. Hild[12], E. Hirose[33], D. Hoak[28], K. A. Hodge[1], K. Holt[28], D. J. Hosken[62], J. Hough[18], D. Hoyland[19], D. Huet[16], B. Hughey[20], S. H. Huttner[18], D. R. Ingram[17], T. Isogai[45], M. Ito[31], A. Ivanov[1], B. Johnson[17], W. W. Johnson[7], D. I. Jones[63], G. Jones[32], R. Jones[18], L. Sancho de la Jordana[64], L. Ju[19], P. Kalmus[1], V. Kalogera[65], S. Kandhasamy[61], J. Kanner[35], D. Kasprzyk[12], E. Katsavounidis[20], K. Kawabe[17], S. Kawamura[66], F. Kawazoe[3], W. Kells[1], D. G. Keppel[1], A. Khalaidovski[3], F. Y. Khalili[26], R. Khan[22], E. Khazanov[67], P. King[1], J. S. Kissel[7], S. Klimenko[10], K. Kokeyama[66], V. Kondrashov[1], R. Kopparapu[51], S. Koranda[4], D. Kozak[1], B. Krishnan[14], A. Kumar[18], P. Kwee[13], P. La Penna[16], P. K. Lam[59], M. Landry[17], B. Lantz[5], M. Laval[9], A. Lazzarini[1], H. Lei[24], M. Lei[1], N. Leindecker[5], I. Leonor[31], N. Leroy[11], N. Letendre[36], C. Li[44], H. Lin[10], P. E. Lindquist[1], T. B. Littenberg[15], N. A. Lockerbie[68], D. Lodhia[12], M. Longo[41], M. Lorenzini[38]‡, V. Loriette[27], M. Lormand[28], G. Losurdo[38]‡, P. Lu[5], M. Lubinski[17], A. Lucianetti[10], H. Lück[3,13], B. Machenschalk[14], M. MacInnis[20], J.-M. Mackowski[56], M. Mageswaran[1], K. Mailand[1], E. Majorana[8]‡, N. Man[9], I. Mandel[65], V. Mandic[61], M. Mantovani[6]||, F. Marchesoni[30]‡||, F. Marion[36], S. Márka[22], Z. Márka[22], A. Markosyan[5], J. Markowitz[20], E. Maros[1], J. Marque[16], F. Martelli[38]‡||, I. W. Martin[18], R. M. Martin[10], J. N. Marx[1], K. Mason[20], A. Masserot[36], F. Matichard[7], L. Matone[22], R. A. Matzner[58], N. Mavalvala[20], R. McCarthy[17], D. E. McClelland[59], S. C. McGuire[53], M. McHugh[69], G. McIntyre[1], D. J. A. McKechan[32], K. McKenzie[59], M. Mehmet[3], A. Melatos[46], A. C. Melissinos[57], G. Mendell[17], D. F. Menéndez[51], F. Menzinger[16], R. A. Mercer[4], S. Meshkov[1], C. Messenger[3], M. S. Meyer[28], C. Michel[56], L. Milano[2]§, J. Miller[18], J. Minelli[51], Y. Minenkov[47]‡, Y. Mino[44], V. P. Mitrofanov[26], G. Mitselmakher[10], R. Mittleman[20], O. Miyakawa[1], B. Moe[4], M. Mohan[16], S. D. Mohanty[24], S. R. P. Mohapatra[37], J. Moreau[27], G. Moreno[17], N. Morgado[56], A. Morgia[47]‡§, T. Morioka[66], K. Mors[3], S. Mosca[2]§, K. Mossavi[3], B. Mours[36], C. MowLowry[59], G. Mueller[10], D. Muhammad[28], H. zur Mühlen[13], S. Mukherjee[24], H. Mukhopadhyay[52], A. Mullavey[59], H. Müller-Ebhardt[3], J. Munch[62], P. G. Murray[18], E. Myers[17], J. Myers[17], T. Nash[1], J. Nelson[18], I. Neri[30]‡§, A. Nishizawa[66], F. Nocera[16], K. Numata[39], E. Ochsner[35], J. O'Dell[34], G. H. Ogin[1], B. O'Reilly[28], R. O'Shaughnessy[51], D. J. Ottaway[62], R. S. Ottens[10], H. Overmier[28], B. J. Owen[51], G. Pagliaroli[47]‡§, C. Palomba[8]‡, Y. Pan[35], C. Pankow[10], F. Paoletti[6]‡,16], M. A. Papa[4,14], V. Parameshwaraiah[17], S. Pardi[2]§, A. Pasqualetti[16], R. Passaquieti[6]§, D. Passuello[6]‡, P. Patel[1], M. Pedraza[1], S. Penn[70], A. Perreca[12], G. Persichetti[2]§, M. Pichot[9], F. Piergiovanni[38]‡||, V. Pierro[41], I. M. Pinto[41], M. Pitkin[18], H. J. Pletsch[3], M. V. Plissi[18], R. Poggiani[6]§, F. Postiglione[49], M. Principe[41], R. Prix[3], G.A. Prodi[54]‡§, L. Prokhorov[26], O. Punken[3], M. Punturo[30]‡, P. Puppo[8]‡, S. van der Putten[23]‡, V. Quetschke[10], F. J. Raab[17], O. Rabaste[21], D. S. Rabeling[23]‡§, H. Radkins[17], P. Raffai[48], Z. Raics[22], N. Rainer[3], M. Rakhmanov[24], P. Rapagnani[8]‡§, V. Raymond[65], V. Re[54]‡§, C. M. Reed[17], T. Reed[71], T. Regimbau[9], H. Rehbein[3], S. Reid[18], D. H. Reitze[10]‡, R. Riesen[28], K. Riles[40], B. Rivera[17], P. Roberts[72], N. A. Robertson[1,18], F. Robinet[11], C. Robinson[32], E. L. Robinson[14], A. Rocchi[47]‡, S. Roddy[28], L. Rolland[36], J. Rollins[22], J. D. Romano[24], R. Romano[2]‡||, J. H. Romie[28], C. Röver[3], S. Rowan[18], A. Rüdiger[3], P. Ruggi[16], P. Russell[1], K. Ryan[17], S. Sakata[66], F. Salemi[54]§, V. Sandberg[17], V. Sannibale[1], L. Santamaría[14], S. Saraf[73], P. Sarin[20], B. Sassolas[56], B. S. Sathyaprakash[32], S. Sato[66], M. Satterthwaite[59], P. R. Saulson[33], R. Savage[17], P. Savov[44], M. Scanlan[71], R. Schilling[3], R. Schnabel[3], R. Schofield[31], B. Schulz[3], B. F. Schutz[14,32], P. Schwinberg[17], J. Scott[18], S. M. Scott[59], A. C. Searle[1], B. Sears[1], F. Seifert[3], D. Sellers[28], A. S. Sengupta[1], D. Sentenac[16], A. Sergeev[67], B. Shapiro[20], P. Shawhan[35], D. H. Shoemaker[20], A. Sibley[28], X. Siemens[4], D. Sigg[17], S. Sinha[5], A. M. Sintes[64], B. J. J. Slagmolen[59], J. Slutsky[7], M. V. van der Sluys[65], J. R. Smith[33], M. R. Smith[1], N. D. Smith[20], K. Somiya[44], B. Sorazu[18],

A. Stein[20], L. C. Stein[20], S. Steplewski[29], A. Stochino[1], R. Stone[24], K. A. Strain[18], S. Strigin[26], A. Stroeer[39], R. Sturani[38‡||], A. L. Stuver[28], T. Z. Summerscales[72], K.-X. Sun[5], M. Sung[7], P. J. Sutton[32], B. L. Swinkels[16], G. P. Szokoly[48], D. Talukder[29], L. Tang[24], D. B. Tanner[10], S. P. Tarabrin[26], J. R. Taylor[3], R. Taylor[1], R. Terenzi[47‡||], J. Thacker[28], K. A. Thorne[28], K. S. Thorne[44], A. Thüring[13], K. V. Tokmakov[18], A. Toncelli[6‡§], M. Tonelli[6‡§], C. Torres[28], C. Torrie[1], E. Tournefier[36], F. Travasso[30‡§], G. Traylor[28], M. Trias[64], J. Trummer[36], D. Ugolini[74], J. Ulmen[5], K. Urbanek[5], H. Vahlbruch[13], G. Vajente[6‡§], M. Vallisneri[44], S. Vass[1], R. Vaulin[4], M. Vavoulidis[11], A. Vecchio[12], G. Vedovato[54||], A. A. van Veggel[18], J. Veitch[12], P. Veitch[62], C. Veltkamp[36], D. Verkindt[36], F. Vetrano[38‡||], A. Viceré[38‡||], A. Villar[1], J.-Y. Vinet[9], H. Vocca[30‡], C. Vorvick[17], S. P. Vyachanin[26], S. J. Waldman[20], L. Wallace[1], H. Ward[18], R. L. Ward[1], M. Was[11], A. Weidner[3], M. Weinert[3], A. J. Weinstein[1], R. Weiss[20], L. Wen[19,44], S. Wen[7], K. Wette[59], J. T. Whelan[14,75], S. E. Whitcomb[1], B. F. Whiting[10], C. Wilkinson[17], P. A. Willems[1], H. R. Williams[51], L. Williams[10], B. Willke[3,13], I. Wilmut[34], L. Winkelmann[3], W. Winkler[3], C. C. Wipf[20], A. G. Wiseman[4], G. Woan[18], R. Wooley[28], J. Worden[17], W. Wu[10], I. Yakushin[28], H. Yamamoto[1], Z. Yan[19], S. Yoshida[76], M. Yvert[36], M. Zanolin[60], J. Zhang[40], L. Zhang[1], C. Zhao[19], N. Zotov[71], M. E. Zucker[20] & J. Zweizig[1]

[1]*LIGO — California Institute of Technology, Pasadena, California 91125, USA. [2]†INFN, ‡Sezione di Napoli; §Università di Napoli 'Federico II' Complesso Universitario di Monte S. Angelo, I-80126 Napoli; ||Università di Salerno, Fisciano, I-84084 Salerno, Italy. [3]*Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, D-30167 Hannover, Germany. [4]*University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA. [5]*Stanford University, Stanford, California 94305, USA. [6]†INFN, ‡Sezione di Pisa; §Università di Pisa, I-56127 Pisa; ||Università di Siena, I-53100 Siena, Italy. [7]*Louisiana State University, Baton Rouge, Louisiana 70803, USA. [8]†INFN, ‡Sezione di Roma; §Università 'La Sapienza', I-00185 Roma, Italy. [9]†Departement Artemis, Observatoire de la Côte d'Azur, CNRS, F-06304 Nice, France. [10]*University of Florida, Gainesville, Florida 32611, USA. [11]†LAL, Université Paris-Sud, IN2P3/CNRS, F-91898 Orsay, France. [12]*University of Birmingham, Birmingham B15 2TT, UK. [13]*Leibniz Universität Hannover, D-30167 Hannover, Germany. [14]*Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, D-14476 Golm, Germany. [15]*Montana State University, Bozeman, Montana 59717, USA. [16]†European Gravitational Observatory (EGO), I-56021 Cascina (Pi), Italy. [17]*LIGO — Hanford Observatory, Richland, Washington 99352, USA. [18]*University of Glasgow, Glasgow G12 8QQ, UK. [19]*University of Western Australia, Crawley, Western Australia 6009, Australia. [20]*LIGO — Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [21]†AstroParticule et Cosmologie (APC), CNRS-UMR 7164-IN2P3-Observatoire de Paris-Université Denis Diderot-Paris VII F-75205 Paris- CEA; DSM/IRSU F-91191 Gif-sur-Yvette, France. [22]*Columbia University, New York, New York 10027, USA. [23]†‡Nikhef, National Institute for Subatomic Physics, P.O. Box 41882, 1009 DB Amsterdam; §The Netherlands VU University Amsterdam, De Boelelaan 1081, 1081 HV, Amsterdam, The Netherlands. [24]*The University of Texas at Brownsville and Texas Southmost College, Brownsville, Texas 78520, USA. [25]*San Jose State University, San Jose, California 95192, USA. [26]*Moscow State University, Moscow 119992, Russia. [27]†ESPCI, CNRS, F-75005 Paris, France. [28]*LIGO — Livingston Observatory, Livingston, Louisiana 70754, USA. [29]*Washington State University, Pullman, Washington 99164, USA. [30]†INFN, ‡Sezione di Perugia; §Università di Perugia, I-6123 Perugia; ||Università di Camerino, I-62032, Camerino, Italy. [31]*University of Oregon, Eugene, Oregon 97403, USA. [32]*Cardiff University, Cardiff CF24 3AA, UK. [33]*Syracuse University, Syracuse, New York 13244, USA. [34]*Rutherford Appleton Laboratory, HSIC, Chilton, Didcot, Oxon OX11 0QX, UK. [35]*University of Maryland, College Park, Maryland 20742, USA. [36]†Laboratoire d'Annecy-le-Vieux de Physique des Particules (LAPP), IN2P3/CNRS, Université de Savoie, F-74941 Annecy-le-Vieux, France. [37]*University of Massachusetts – Amherst, Amherst, Massachusetts 01003, USA. [38]†INFN, ‡Sezione di Firenze, I-50019 Sesto Fiorentino; §Università degli Studi di Firenze, I-50121, Firenze; ||Università degli Studi di Urbino 'Carlo Bo', I-61029 Urbino, Italy. [39]*NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [40]*University of Michigan, Ann Arbor, Michigan 48109, USA. [41]*University of Sannio at Benevento, I-82100 Benevento, Italy. [42]*The University of Mississippi, University, Mississippi 38677, USA. [43]*Charles Sturt University, Wagga Wagga, New South Wales 2678, Australia. [44]*Caltech-CaRT, Pasadena, California 91125, USA. [45]*Carleton College, Northfield, Minnesota 55057, USA. [46]*The University of Melbourne, Parkville, Victoria 3010, Australia. [47]†INFN, ‡Sezione di Roma Tor Vergata; §Università di Roma Tor Vergata; ||Istituto di Fisica dello Spazio Interplanetario (IFSI) INAF, I-00133 Roma; ¶Università dell'Aquila, I-67100 L'Aquila, Italy. [48]*Eötvös University, ELTE 1053 Budapest, Hungary. [49]*University of Salerno, 84084 Fisciano (Salerno), Italy. [50]*The University of Sheffield, Sheffield S10 2TN, UK. [51]*The Pennsylvania State University, University Park, Pennsylvania 16802, USA. [52]*Inter-University Centre for Astronomy and Astrophysics, Pune 411007, India. [53]*Southern University and A&M College, Baton Rouge, Louisiana 70813, USA. [54]†INFN, ‡Gruppo Collegato di Trento and §Università di Trento, I-38050 Povo, Trento, Italy; INFN, ||Sezione di Padova and ¶Università di Padova, I-35131 Padova, Italy. [55]*California Institute of Technology, Pasadena, California 91125, USA. [56]†Laboratoire des Matériaux Avancés (LMA), IN2P3/CNRS, F-69622 Villeurbanne, Lyon, France. [57]*University of Rochester, Rochester, New York 14627, USA. [58]*The University of Texas at Austin, Austin, Texas 78712, USA. [59]*Australian National University, Canberra 0200, Australia. [60]*Embry-Riddle Aeronautical University, Prescott, Arizona 86301, USA. [61]*University of Minnesota, Minneapolis, Minnesota 55455, USA. [62]*University of Adelaide, Adelaide, South Australia 5005, Australia. [63]*University of Southampton, Southampton SO17 1BJ, UK. [64]*Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain. [65]*Northwestern University, Evanston, Illinois 60208, USA. [66]*National Astronomical Observatory of Japan, Tokyo 181-8588, Japan. [67]*Institute of Applied Physics, Nizhny Novgorod, 603950, Russia. [68]*University of Strathclyde, Glasgow G11XQ, UK. [69]*Loyola University, New Orleans, Louisiana 70118, USA. [70]*Hobart and William Smith Colleges, Geneva, New York 14456, USA. [71]*Louisiana Tech University, Ruston, Louisiana 71272, USA. [72]*Andrews University, Berrien Springs, Michigan 49104, USA. [73]*Sonoma State University, Rohnert Park, California 94928, USA. [74]*Trinity University, San Antonio, Texas 78212, USA. [75]*Rochester Institute of Technology, Rochester, New York 14623, USA. [76]*Southeastern Louisiana University, Hammond, Louisiana 70402, USA.

*The LIGO Scientific Collaboration.

†The Virgo Collaboration.

LETTERS

In situ observation of incompressible Mott-insulating domains in ultracold atomic gases

Nathan Gemelke[1], Xibo Zhang[1], Chen-Lung Hung[1] & Cheng Chin[1]

The observation of the superfluid to Mott insulator phase transition of ultracold atoms in optical lattices[1] was an enabling discovery in experimental many-body physics, providing the first tangible example of a quantum phase transition (one that occurs even at zero temperature) in an ultracold atomic gas. For a trapped gas, the spatially varying local chemical potential gives rise to multiple quantum phases within a single sample, complicating the interpretation of bulk measurements[1–5]. Here we report spatially resolved, *in-situ* imaging of a two-dimensional ultracold atomic gas as it crosses the superfluid to Mott insulator transition, providing direct access to individual characteristics of the insulating, superfluid and normal phases. We present results for the local compressibility in all phases, observing a strong suppression in the insulator domain and suppressed density fluctuations for the Mott insulator, in accordance with the fluctuation–dissipation theorem. Furthermore, we obtain a direct measure of the finite temperature of the system. Taken together, these methods enable a complete characterization of multiple phases in a strongly correlated Bose gas, and of the interplay between quantum and thermal fluctuations in the quantum critical regime.

Since its theoretical inception[6–8], two of the most celebrated properties of the bosonic Mott insulator have been its incompressibility and suppression of local density fluctuations[9], induced by the enhanced importance of inter-particle repulsion for particles subject to a strong lattice potential. The result for a trapped atom gas is the remarkable 'wedding-cake' density profile, in which successive Mott-insulator domains appear as plateaus of constant density. Related phenomena have been studied through the coherence[1,2], transport[1,3], noise correlations[4], and number variance[5,10], but direct observation of the incompressibility has proved difficult owing to the inhomogeneous nature of all experiments to date, and to the technical difficulty of making spatially resolved measurements. Innovative experimental efforts incorporating tomographic imaging and other advanced techniques have yielded evidence[11,12] that shell structure exists in the Mott insulator regime, though no-one has directly observed the incompressibility of the insulating density plateaus by imaging a complete and single physical system *in situ*.

We report studies based on direct *in-situ* imaging of an atomic Mott insulator. By loading a degenerate Bose gas of caesium-133 atoms into a thin layer of a two-dimensional (2D) optical lattice potential, and adiabatically increasing the optical lattice depth, we observe the emergence of an extremely flat density near the centre of the cloud, which corresponds to a Mott-insulator phase with accurately one atom per site. From density profiles, we extract important thermodynamic and statistical information, confirming the incompressibility and reduction of density fluctuations in the Mott insulator as described by the fluctuation–dissipation theorem.

The single-layer 2D optical lattice is formed by two pairs of counter-propagating laser beams derived from a Yb fibre laser at wavelength

$\lambda = 1,064$ nm. The pairs are oriented orthogonally on the horizontal (x–y) plane, forming a square lattice with site spacing $d = \lambda/2 = 0.532\,\mu\text{m}$. A weak harmonic potential of $V_h = m(\omega_x^2 x^2 + \omega_y^2 y^2)/2$ localizes the sample, where m is the caesium mass, and the geometric mean of the principal trap frequencies $\omega_{x,y}$ is $\omega_r = \sqrt{\omega_x \omega_y} = 2\pi \times 9.5$ Hz, where subscript r refers to radial direction (a weak dependence on lattice depth is described in the Methods). Vertical confinement is provided by an additional vertical optical lattice with a site spacing of $4\,\mu\text{m}$, formed by two beams intersecting at an angle of $15°$, confining atoms in a Gaussian wavepacket of width (oscillator length) $a_z = 0.30\,\mu\text{m}$. The sample is loaded into a single site of the vertical lattice, kept deep to prevent vertical tunnelling. Tunnelling in the horizontal 2D lattice is controlled by varying the lattice depth V (ref. 1). Details on preparation of the atomic sample can be found in the Supplementary Information and ref. 13.

We obtain a top view of the sample using absorption imaging, directly revealing the atomic surface density $n(x, y)$ on the horizontal plane. The imaging resolution is $3–4\,\mu\text{m}$, and magnification such that one imaging pixel corresponds to an area of $(2\,\mu\text{m})^2$ on the object plane. Unit filling in a 2D optical lattice has a conveniently measurable optical absorption on resonance.

The superfluid to Mott insulator transition of ultracold atoms in an optical lattice is described by the Bose–Hubbard model, characterized by
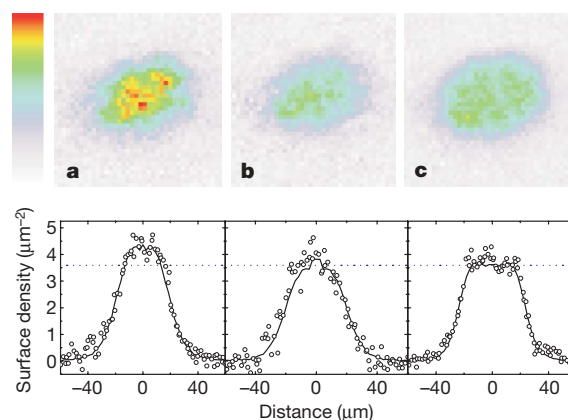
Figure 1 | False-colour absorption images and line cuts along major axis of density profiles for $N = 7,500$ ultracold caesium atoms at scattering length $a = 310a_B$ in 2D optical lattices. a, Superfluid regime (shallow lattice $V = 2.4E_R$); **b**, phase transition regime (medium lattice depth $V = 9.4E_R$); and **c**, Mott-insulator regime (deep lattice $V = 22E_R$). Images are averaged over three experiment repetitions. The colour scale shows linear variation with density from zero to the peak value of $5.4\,\mu\text{m}^{-2}$. Line cuts are taken along the major axis, and compared to radial average of density (solid line) over the entire image as described in text. The blue horizontal dotted line indicates the density of one atom per site.

[1]The James Franck Institute and Department of Physics, University of Chicago, Chicago, Illinois 60637, USA.
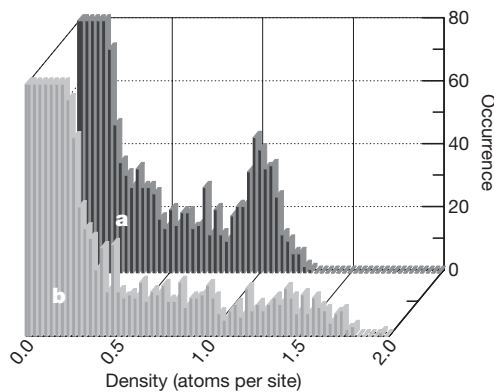
on-site interaction U and the tunnelling strength J (ref. 8). In 2D optical lattices, superfluid is converted into a Mott insulator when U/J exceeds 16 near the density of one atom per site[2,14]. Here, the superfluid–Mott insulator phase transition can be induced by either increasing the lattice potential depth V (typically measured in units of recoil energy $E_R = h \times 1.3$ kHz, where h is Planck's constant)[1,2,12] or the atomic interaction strength (characterized by scattering length a, typically measured in Bohr radii a_B) via a magnetically tuned Feshbach resonance[15], together providing complete, independent control of U and J.

Atomic density profiles in the lattice are shown in Fig. 1. For weak lattice depths (superfluid regime), the density profiles are bell-shaped, with negative curvature at the centre (Fig. 1a), indicating a finite, positive compressibility dictated by the interaction coupling constant (discussed below). In sufficiently deep lattices, we observe a flattened density at the centre of the sample (Fig. 1b, c), indicating development of a Mott-insulating phase with one particle per lattice site. This density plateau, an important feature of the Mott-insulator phase, arises from incompressibility.

A primary check on the Mott insulator is to compare the measured density in the plateau to that corresponding to one atom per site, given by Mott-insulator physics as a 'standard candle' of atomic density. Using the known scattering cross-section, correcting for saturation effects (see Methods), we determine the plateau density (and standard error, in parentheses) to be $n = 3.5(3)$ μm^{-2}, in agreement with the expected value $1/d^2 = 3.53$ μm^{-2}.

To distinguish a Mott insulator from superfluid or normal gas, we histogram the occurrence of pixels $h(n)$ in the images corresponding to a density n within a small bin size Δn. The Mott-insulator plateau, containing a large number of pixels with similar atomic density, appears as a peak at $n = 1/d^2$ (Fig. 2a). In general, the occurrence of a particular density n can be regarded as the rate at which local chemical potential changes with density, multiplied by the number of pixels $w(\mu)\Delta\mu$ corresponding to a chemical potential between μ and $\mu + \Delta\mu$. The occurrence at density n is then $h(n) = \Delta n \omega(\mu)\Delta\mu/\Delta n \approx \Delta n \omega(\mu)\kappa^{-1}$, where $\kappa = \partial n/\partial\mu$ is the local compressibility[16]. In a harmonic trap, $\omega(\mu) = 2\pi/md^2\omega_r^2$ is constant, and the histogram is a particularly useful tool for distinguishing different phases. For a pure Bose–Einstein condensate (BEC) in the Thomas–Fermi limit, the compressibility is constant up to the peak density n_{peak}, and results in a constant $h(n)$ for $n \leq n_{peak}$ (see Fig. 2b for $0.5/d^2 < n < 1.5/d^2$). For the Mott insulator, the density is insensitive to chemical potential in a narrow range near $n = 1/d^2$, indicating a vanishing compressibility, and thus a sharp histogram peak at $n = 1/d^2$. The peak's presence in Fig. 2a is thus directly related to the incompressibility in the Mott phase. Finally, the compressibility of a normal (ideal) gas is proportional to its density, so $h(n) \propto 1/n$, leading to the strong upturn at low densities in Fig. 2a, b for both regimes.

Much more information can be obtained from the density profiles, as recently suggested in ref. 17. For example, the compressibility in a 2D cylindrically symmetric trap can be written $\kappa = \partial n/\partial\mu = -n'(r)/(rm\omega_r^2)$, where we have assumed the local density approximation, and that the chemical potential depends on the trapping potential $\mu = \mu_0 - V_h(r)$. For a BEC in the Thomas–Fermi regime, the compressibility is positive and constant, $\kappa_{BEC} = 1/g$, where $g = \sqrt{8\pi}a\hbar^2/ma_z$ is the 2D interaction parameter[18]. We can thereby relate the measured compressibility to that of a BEC as:

$$\frac{\kappa}{\kappa_{BEC}} = -\left(\frac{2}{\pi}\right)^{7/2} \frac{n'(r)}{rd^{-4}} \frac{a}{a_z} \left(\frac{E_R}{\hbar\omega_r}\right)^2 \qquad (1)$$

We evaluate κ from azimuthally averaged density profiles (Fig. 3a). The eccentricity of the trap is corrected by rescaling the principal axes as determined from the density profile, and verified to be consistent with direct measurement of trap frequencies. Owing to the singular nature of $n'(r)/r$ near the centre, we evaluate κ there by fitting $n(r)$ to a quadratic, $n(r) = n(0) - \alpha r^2$. The curvature α then gives the compressibility as $\kappa(0) = 2\alpha/m\omega_r^2$, for which we obtain $\kappa/\kappa_{BEC} = 0.34(10)$ in a weak lattice and $\kappa/\kappa_{BEC} = 0.013(6)$ in a strong lattice (see Fig. 3). In the weak lattice (superfluid regime), the finite and constant compressibility at the centre agrees with expectation for the superfluid phase, though lower than expected, which we attribute to finite temperature and calibration of trap parameters. The finite temperature is also clear
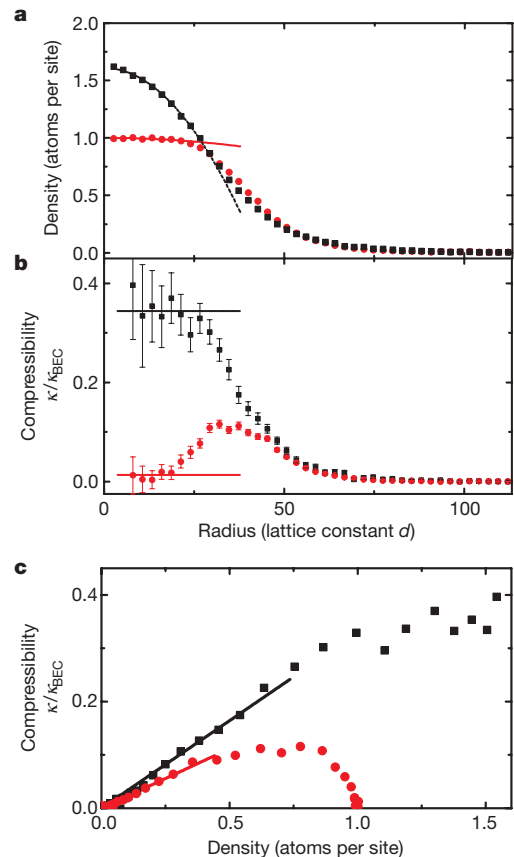
Figure 3 | Extraction of compressibility from density profiles. a, Radially averaged profiles (three images) in the superfluid (black squares, $V = 0.3E_R$, $N = 7,200$) and Mott insulator (red circles, $V = 22E_R$, $N = 6,700$), with $a = 310a_B$. A quadratic fit to the sample's centre extracts the curvature near $r = 0$. **b,** Normalized compressibilities derived from **a** using equation (1) in the superfluid (black squares) and Mott-insulator (red circles) regimes. The horizontal lines indicate compressibility near $r = 0$, estimated from the quadratic fits in **a**. Rising compressibility at $r = 20d$ marks the Mott-insulator boundary. **c,** The dependence of compressibility on atomic density. Linear dependence at low densities (normal gas) is best-fitted by solid lines. Error bars indicate standard error in the mean.

Figure 2 | Histograms of density profiles in the Mott-insulator regime and the superfluid regime. a, Mott-insulator regime: $V = 38E_R$, $a = 460a_B$. **b,** Superfluid regime: $V = 0.5E_R$, $a = 460a_B$. The histograms are based on an average of three density images. The bin size is $\Delta n = 0.03$. Occurrence is the number of samples (pixels) measured as a density between n and $n + \Delta n$.

in the exponential tail of the density profile and the compressibility[19], from which we derive the temperature 10(2) nK in the superfluid regime ($V = 0.3E_R$) and 15(3) nK in the Mott-insulator regime ($V = 22E_R$).
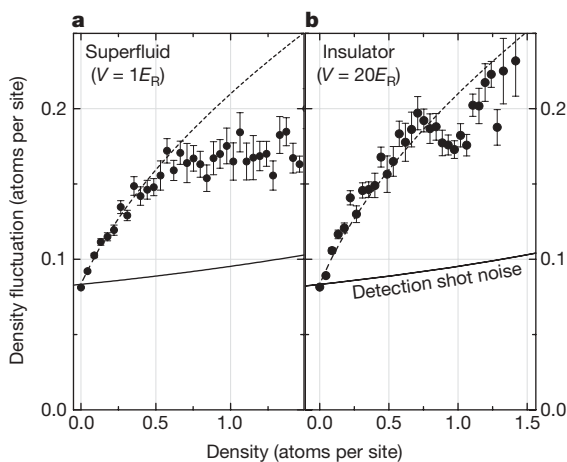
In a deep lattice (Mott-insulator regime), we observe a strong reduction of the compressibility in the trap centre, below that in the superfluid phase for the weak lattice, strongly supporting the emergence of a Mott-insulator phase at the centre of the sample. Away from the centre, κ suddenly increases at $r = 20d$, then decreases for $r > 40d$. The exponential decay is again consistent with a normal gas. Between a Mott insulator and a normal gas ($20d < r < 40d$), a more detailed measurement and model of compressibility would be necessary to identify the local phase.

Within the local density approximation, one may consider any small area of the sample as a thermodynamic subsystem in a grand-canonical ensemble, assumed to be in equilibrium with the remainder of the gas. One can then invoke the fluctuation–dissipation theorem (see for example, refs 16 and 20) to ascertain that incompressibility necessarily implies a low local particle number fluctuation; this relationship takes the form:

$$\delta n^2 \approx \kappa k_B T \qquad (2)$$

Resolved *in-situ* imaging provides an enticing opportunity to measure fluctuations of the local density[9,21], and thus check the validity of the fluctuation–dissipation theorem. We measured fluctuations by recording multiple absorption images, calculating the variance of density measured in each pixel (each collects signals from a patch of $(2\,\mu m/d)^2 \approx 14$ lattice sites). Figure 4 shows the recorded fluctuations, where pixels are binned according to their mean atomic density. Fluctuations consist of detection (photo-electron shot) noise and thermal and quantum atomic density fluctuations. Detection shot noise can be well-calibrated and modelled by analysing portions of the images with low density; extension to higher optical depth (density) shows the weak dependence illustrated in Fig. 4.

Above the detection noise, density fluctuations (see Fig. 4) show a strong qualitative agreement with the compressibility presented in Fig. 3, as expected from the fluctuation–dissipation theorem. For example, the Mott-insulator phase shows a strong suppression of fluctuations at the density of one atom per site. The superfluid regime lacks this feature, instead showing a pronounced flattening as the

sample transitions from normal gas to superfluid, as expected from the constant compressibility in the superfluid phase (Fig. 3c). Finally, at low density, the normal gas shows a temperature-independent fluctuation of $\delta n = \gamma \sqrt{n}$, which can be anticipated from Fig. 3c, and agrees with the fluctuation–dissipation theorem. The coefficient γ is roughly consistent with the fluctuation–dissipation theorem, and the measured imaging resolution (see Methods).

Clearly, *in-situ* imaging of the Mott insulator is a powerful new tool with which to investigate new quantum phases of cold atoms in optical lattices. From the density profiles, not only can we observe the density plateau, incompressibility and reduction of fluctuations in the Mott insulating phase, but also demonstrate a qualitative validation of the fluctuation–dissipation theorem. Relatively modest extension of this work holds new promise for studying the role of quantum fluctuations, correlation and thermodynamics near a quantum phase transition.

METHODS SUMMARY

Caesium Bose condensates are produced by forced evaporative cooling in a crossed-beam dipole trap. The condensate is compressed vertically by loading into a single layer of an optical lattice with the scattering length tuned near zero. After this, the lattice is adiabatically applied by controlled retroreflection of dipole trapping beams, and the scattering length brought to its final value. Imaging is performed absorptively along the vertical, calibrated for saturation effects by varying the intensity of the imaging light. Fluctuations of density are calculated for each pixel in a series of images taken at identical experimental parameters, and plotted against the mean density at that pixel. The parameter γ is estimated from a model of the expected averaging of thermal fluctuations over an imaging resolution limited spot. Further details may be found in the online-only Methods and in the Supplementary figure.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Greiner, M., Mandel, O., Esslinger, T., Hansch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
2. Spielman, I. B., Phillips, W. D. & Porto, J. V. Condensate fraction in a 2D Bose gas measured across the Mott-insulator transition. *Phys. Rev. Lett.* **100**, 120402 (2008).
3. Kohl, M., Moritz, H., Stoferle, T., Schori, C. & Esslinger, T. Superfluid to Mott insulator transition in one, two and three dimensions. *J. Low-Temp. Phys.* **138**, 635–644 (2005).
4. Folling, S. *et al.* Spatial quantum noise interferometry in expanding ultracold atom clouds. *Nature* **434**, 481–484 (2005).
5. Gerbier, F., Folling, S., Widera, A., Mandel, O. & Bloch, I. Probing number squeezing of ultracold atoms across the superfluid-Mott insulator transition. *Phys. Rev. Lett.* **96**, 090401 (2006).
6. Kaganov, M. I. & Chubukov, A. V. Interacting magnons. *Uspekhi Fizicheskikh Nauk.* **153**, 537–578 (1987).
7. Fisher, M. P. A., Weichman, P. B., Grinstein, G. & Fisher, D. S. Boson localization and the superfluid-insulator transition. *Phys. Rev. B* **40**, 546–570 (1989).
8. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
9. Capogrosso-Sansone, B., Prokof'ev, N. V. & Svistunov, B. V. Phase diagram and thermody-namics of the three-dimensional Bose-Hubbard model. *Phys. Rev. B* **75**, 134302 (2007).
10. Greiner, M., Mandel, O., Hansch, T. W. & Bloch, I. Collapse and revival of the matter wave field of a Bose-Einstein condensate. *Nature* **419**, 51–54 (2002).
11. Folling, S., Widera, A., Muller, T., Gerbier, F. & Bloch, I. Formation of spatial shell structure in the superfluid to Mott insulator transition. *Phys. Rev. Lett.* **97**, 060403 (2006).
12. Campbell, G. K. *et al.* Imaging the Mott insulator shells by using atomic clock shifts. *Science* **313**, 649–652 (2006).
13. Hung, C. L., Zhang, X., Gemelke, N. & Chin, C. Accelerating evaporative cooling of atoms into Bose-Einstein condensation in optical traps. *Phys. Rev. A* **78**, 011604 (2008).
14. Spielman, I. B., Phillips, W. D. & Porto, J. V. Mott-insulator transition in a two-dimensional atomic Bose gas. *Phys. Rev. Lett.* **98**, 080404 (2007).
15. Chin, C., Grimm, R., Julienne, P. & Tiesinga, E. Feshbach resonances in ultracold gases. Preprint at ⟨http://arxiv.org/abs/0812.1496⟩ (2008).
16. Batrouni, G. G. *et al.* Mott domains of bosons confined on optical lattices. *Phys. Rev. Lett.* **89**, 117203 (2002).
17. Ho, T. -L. & Zhou, Q. Obtaining phase diagram and thermodynamic quantities of bulk systems from the densities of trapped gases. Preprint at ⟨http://arxiv.org/abs/0901.0018⟩ (2008).

Figure 4 | The fluctuation of local density extracted from a set of eleven absorption images in the weak and deep lattice regimes. The insulator and superfluid show a pronounced difference at one atom per site, where the insulator's fluctuation is suppressed by incompressibility. In the superfluid, constant compressibility initiates a flattening. At low densities, in both the weak (**a**) and the deep (**b**) lattice regimes, the fluctuation shows a characteristic $\sqrt{n}$ dependence, where the gas is presumed to be normal; the dashed line shows the best-fitted $\sqrt{n}$ dependence. The total number of atoms was $N = 8,300$ (superfluid) and $N = 9,600$ (Mott insulator) with $a = 310a_B$ for both sets. Error bars indicate standard error in the mean.

18. Tanatar, B., Minguzzi, A., Vignolo, P. & Tosi, M. P. Density profile of a Bose-Einstein condensate inside a pancake-shaped trap: observational consequences of the dimensional cross-over in the scattering properties. *Phys. Lett. A* **302**, 131–136 (2002).

19. Gerbier, F. Boson Mott insulators at finite temperatures. *Phys. Rev. Lett.* **99**, 120405 (2007).

20. Huang, K. *Statistical Mechanics* 152–154 (Wiley, 1963).

21. Esteve, J. *et al.* Observations of density fluctuations in an elongated Bose gas: ideal gas and quasicondensate regimes. *Phys. Rev. Lett.* **96**, 090401 (2006).

Author Contributions All authors contributed to the analysis and writing of this manuscript; construction of the apparatus and acquisition of data was primarily the responsibility of C-L.H. and X.Z.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.C. (cchin@uchicago.edu).

METHODS

Preparation of BEC in a thin 2D optical lattice. The ^{133}Cs BEC is formed in a crossed-beam dipole trap by an efficient evaporative cooling method[13]. The dipole trap consists of three beams on the horizontal plane (see Supplementary figure): two orthogonal beams at the wavelength of 1,064 nm (Yb fibre laser, YLR-20-1064-LP-SF, IPG Photonics), focused to $1/e^2$ radii of 350 μm, and one CO_2 laser beam (not shown in the Supplementary figure) at the wavelength of 10.6 μm (Gem-Select 100, Coherent), focused to a vertical $1/e^2$ radius of 70 μm and horizontal radius of 2 mm. The CO_2 beam intersects the Yb fibre laser beams at an angle of 45° and provides an enhanced vertical confinement to support the atoms against gravity. With $N = 10^4$ atoms in a pure condensate, the Thomas–Fermi radii of the condensate are $(r_x, r_y, r_z) = (23, 14, 3.6)$ μm.

After a pure BEC is obtained, the sample is compressed vertically by introducing a vertical lattice, formed by two laser beams (Mephisto, Innolight) inclined at $+7.5°$ and $−7.5°$ relative to the horizontal plane. The vertical lattice has a spacing of 4 μm and, together with the crossed dipole trap, forms an array of 2D oblate 'pancake' potentials, with harmonic confinement frequencies of 850 Hz at its maximum depth.

To load the condensate into a single pancake trap, we first ramp the magnetic field to 17.2 G in 400 ms, reducing the s-wave scattering length to $a < 10a_B$, and then turn on the vertical lattice in 100 ms. Atomic population in other lattice sites, if any, can be identified by observing an interference pattern in time-of-flight images taken from the side. For this work, we observe a sufficiently weak interference pattern contrast, and hence conclude that over 98% of the atoms are in a single pancake trap. After the vertical lattice is fully turned on, the CO_2 laser intensity is ramped to zero in 100 ms while the scattering length is ramped to its final value by tuning the external magnetic field.

The 2D lattice potential in the horizontal (x- and y-) directions is formed by introducing retro-reflections of the 1,064 nm dipole trap beams. A continuous evolution from a pure dipole trap (with zero retro-reflection) to a 2D optical lattice (with significant retro-reflection) is achieved by passing each dipole trap beam (after it passes through the atomic cloud once) through two acousto-optic modulators controlled by the same radio-frequency source, then off a retroreflection mirror. The acousto-optic modulators induce an overall zero-frequency shift, but permit a dynamic control of the retroreflection intensity over six orders of magnitude. To load the lattice to a depth of $38E_r$, the retroreflection intensities are slowly ramped over 200 ms with an exponential waveform of 36 ms time constant. For smaller final lattice depths, the ramp waveform is fixed but duration shortened. Onsite interaction energy U and tunnelling rate J are evaluated from the measurements of the lattice vibration frequencies and band structure calculation. Envelope trapping frequencies were separately measured by exciting centre-of-mass oscillations, and found to be consistent with orientation and ellipticity of

in-situ images of atomic density. A weak variation of the mean envelope frequency with lattice depth was measured and accounted for by the expression: $\omega_r = \sqrt{\omega_x \omega_y} = 2\pi \times 9.5(1 + V/82E_R)$ Hz.

Calibration of atomic surface density. By varying the intensity of the imaging beam, we measure the optical depth on resonance in the density plateau using $OD = \ln(M_0/M)$, where M is the number of photons collected by a charge-coupled device (CCD) pixel in the presence of the atoms and M_0 is that without the atoms. The optical depth in the plateau is extracted from a fit to the peak in the histogram. We then assume $OD = n\sigma/(1 + M_0/M_{sat})$ and fit the variation of peak optical depth to determine the optical depth in the zero intensity limit $M_0 \to 0$, and thus the surface density of the sample. Here, $\sigma = 0.347$ μm^2 is the known caesium-atom optical cross-section and the fit parameter M_{sat} represents the photon number on a CCD pixel at the atomic saturation intensity.

Fluctuation of atomic density. The fluctuations in the absorption images are estimated by taking the average of 11 images under the same experimental procedure, and calculating the mean and variance of optical depth measured at each CCD pixel. Fluctuations are presumed to arise from optical shot noise, thermal atomic fluctuation, and long-lengthscale variations arising from total atom number fluctuation. The optical shot noise is calibrated by examining regions with negligible atomic density, and extended to higher optical depth using $\delta OD_{\text{optical shot}} \propto \sqrt{1 + e^{OD}}$. For the thermal cloud, with density $n < 0.3$ atoms per site, the fluctuation–dissipation theorem predicts $\delta N_a = \sqrt{N_a}$, with N_a the number of atoms measured in a given region. This result should be valid for a region significantly larger than the correlation length, which we expect for the normal gas to be of the order of the de Broglie thermal wavelength, expected to be less than 1.5 μm for our sample. Although each imaging pixel corresponds to an area in the object plane consisting of ~14 sites, imperfect imaging resolution is expected to effectively average away a certain fraction of the total fluctuation. This effect can be calculated, assuming statistical independence for each site, by summing the weight $w_{i,j}$ of a resolution-limited spot falling within a given pixel j for each lattice site i, giving a variance reduced by $\sum_i w_{i,j}^2$. The result for our parameters is a reduction to $\delta n = \gamma \sqrt{n}$, with $\gamma \approx 0.11(1)$. This should be compared with the fraction of the total fluctuation shown in Fig. 4 corresponding to thermal fluctuations in the superfluid regime. To make this comparison, we reject global fluctuations associated with variation of the total atom number by subtracting the variance we calculate after first applying a resolution-spoiling Gaussian blur to the images from the variance without modification. We find, for the remaining high-spatial-frequency fluctuations, a best fit to γ of 0.15(2), using a Gaussian blur $1/e^2$ radius of $r_b = 14$ μm to remove global variations (the result varies within stated error for blur radii 7 μm $< r_b <$ 28 μm). The remaining discrepancy is probably due to calibration of imaging resolution, and possibly the effect of a non-negligible correlation length.

Satellite-based estimates of groundwater depletion in India

Matthew Rodell[1], Isabella Velicogna[2,3,4] & James S. Famiglietti[2]

Groundwater is a primary source of fresh water in many parts of the world. Some regions are becoming overly dependent on it, consuming groundwater faster than it is naturally replenished and causing water tables to decline unremittingly[1]. Indirect evidence suggests that this is the case in northwest India[2], but there has been no regional assessment of the rate of groundwater depletion. Here we use terrestrial water storage-change observations from the NASA Gravity Recovery and Climate Experiment satellites[3] and simulated soil-water variations from a data-integrating hydrological modelling system[4] to show that groundwater is being depleted at a mean rate of $4.0 \pm 1.0\,\mathrm{cm\,yr^{-1}}$ equivalent height of water ($17.7 \pm 4.5\,\mathrm{km^3\,yr^{-1}}$) over the Indian states of Rajasthan, Punjab and Haryana (including Delhi). During our study period of August 2002 to October 2008, groundwater depletion was equivalent to a net loss of $109\,\mathrm{km^3}$ of water, which is double the capacity of India's largest surface-water reservoir. Annual rainfall was close to normal throughout the period and we demonstrate that the other terrestrial water storage components (soil moisture, surface waters, snow, glaciers and biomass) did not contribute significantly to the observed decline in total water levels. Although our observational record is brief, the available evidence suggests that unsustainable consumption of groundwater for irrigation and other anthropogenic uses is likely to be the cause. If measures are not taken soon to ensure sustainable groundwater usage, the consequences for the 114,000,000 residents of the region may include a reduction of agricultural output and shortages of potable water, leading to extensive socio-economic stresses.

Groundwater responds more slowly to meteorological conditions than the near-surface components of the terrestrial water cycle[5]. Its residence time (the ratio of quantity in storage to average rate of recharge or discharge) ranges from months in shallow aquifers to a million or more years in deep desert aquifers[6]. Hence, groundwater can be slow to recover from perturbations to its state of dynamic equilibrium. In particular, withdrawals can easily surpass net recharge in arid and semi-arid regions where people depend on fresh groundwater for domestic needs and irrigation[1]. Despite the increasing pressure placed on water resources by population growth and economic development, the laws governing groundwater rights have not changed accordingly, even in developed nations[7]. Nor is groundwater depletion limited to dry climates: pollution and mismanagement of surface waters can cause over-reliance on groundwater in regions where annual rainfall is abundant.

India now suffers severe water shortages in many of its states. It averages about $120\,\mathrm{cm\,yr^{-1}}$ of precipitation, which is more than any other country of comparable size[8], but the rain is unevenly distributed. In New Delhi, India's richest city, most middle-class residents do not have a dependable source of clean water (Sengupta, S., 'In

India, water crisis means foul sludge', New York Times, 29 September 2006). The World Bank has warned that India is on the brink of a severe water crisis[9]. Nationally, groundwater accounts for about 50–80% of domestic water use and 45–50% of irrigation[8,10]. Total irrigated area in India nearly tripled to 33,100,000 ha between 1970 and 1999[11]. In neighbouring Pakistan, which is largely arid, groundwater is essential for much of the country's agriculture. Competition for precious water in transboundary aquifers is likely to exacerbate already strained relations between the two nations.

India's government is aware that groundwater is being withdrawn at unsustainable rates in some areas, and in 1986 it established a Central Ground Water Authority with the power to regulate groundwater development[12]. However, as in other nations composed of smaller sovereignties and encompassing competing interests that have become dependent on a certain level of water availability, it is difficult to implement a coordinated and appropriately stringent response. Political and aquifer boundaries bear no resemblance to each other, and aquifers themselves are interconnected, so that one state's (or country's) groundwater management practices are likely to affect its neighbour. Holistic regional groundwater assessments would be valuable in promoting appropriate policies and for hydrologic research, but such assessments are difficult to generate on the basis of well surveys, which are typically unsystematic.

The Gravity Recovery and Climate Experiment (GRACE) satellite mission, launched by NASA and the German Aerospace Centre (DLR) in 2002, measures temporal variations in the gravity field, which can be used to estimate changes in terrestrial water storage (TWS)[3]. Although its spatial resolution (no better than $\sim 160,000\,\mathrm{km^2}$) and temporal resolution (ten day to monthly) are low in comparison with those of other Earth-observing satellites, GRACE has the major advantage that it senses water stored at all levels, including groundwater. Unlike radars and radiometers, it is not limited to measurement of atmospheric and near-surface phenomena.

Groundwater storage variations can be isolated from GRACE data given auxiliary information on the other components of TWS, from either *in situ* observations[13] or land-surface models[14]. We used the second approach to produce a time series of groundwater storage anomalies (deviations from the mean state) averaged over the area encompassed by Rajasthan ($342,239\,\mathrm{km^2}$), Punjab ($50,362\,\mathrm{km^2}$) and Haryana ($45,695\,\mathrm{km^2}$ including the National Capital Territory of Delhi) between August 2002 and October 2008. This region was chosen because the Indian Ministry of Water Resources estimates that groundwater withdrawals in each of the three states exceed recharge[2] (Fig. 1). Figure 2 maps the averaging function used to retrieve regional TWS time series from the GRACE data.

Rajasthan, Punjab and Haryana are semi-arid to arid, averaging about 50 cm of annual rainfall overall[15–17], and encompass the eastern part of the Thar Desert. The 114,000,000 residents of the region have

[1]Hydrological Sciences Branch, Code 614.3, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [2]Department of Earth System Science, University of California, Irvine, California 92697-3100, USA. [3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109-8099, USA, [4]Department of Physics, University of Udine, 208 Via delle Scienze, 33100 Udine, Italy.
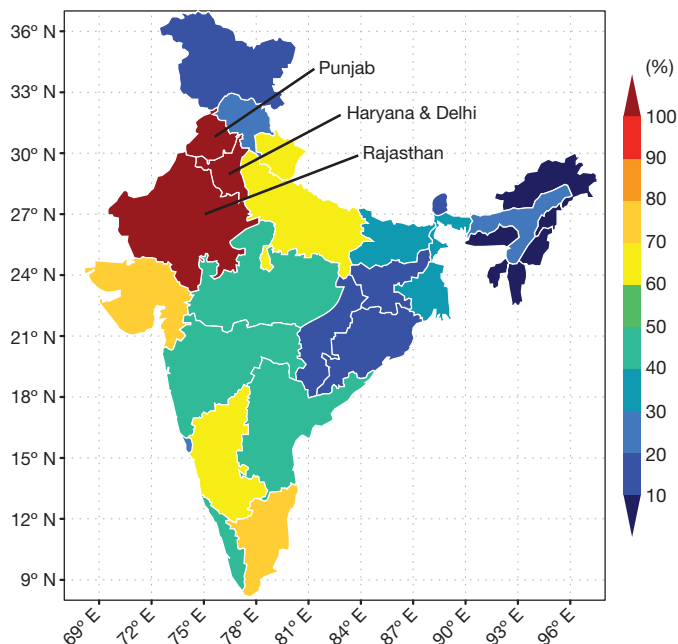
Figure 1 | Groundwater withdrawals as a percentage of recharge. The map is based on state-level estimates of annual withdrawals and recharge reported by the Indian Ministry of Water Resources[2]. The three states studied here are labelled.

benefitted from India's 'green revolution', a massive agricultural expansion fuelled largely by increased production of groundwater for irrigation, which began in the 1960s. Wheat, rice and barley are the major crops. The region is underlain by the Indus River plain aquifer, a 560,000 km[2] unconfined-to-semiconfined porous alluvial formation that straddles the border between India and Pakistan[11].

TWS variations observed by GRACE include the combined contributions of groundwater, soil water, surface water (lakes, rivers, canals and rice paddies), snow, ice and biomass. Interannual variations in biomass have been shown to be well below the detection limits of GRACE[18]. The climate of the study region is warm and
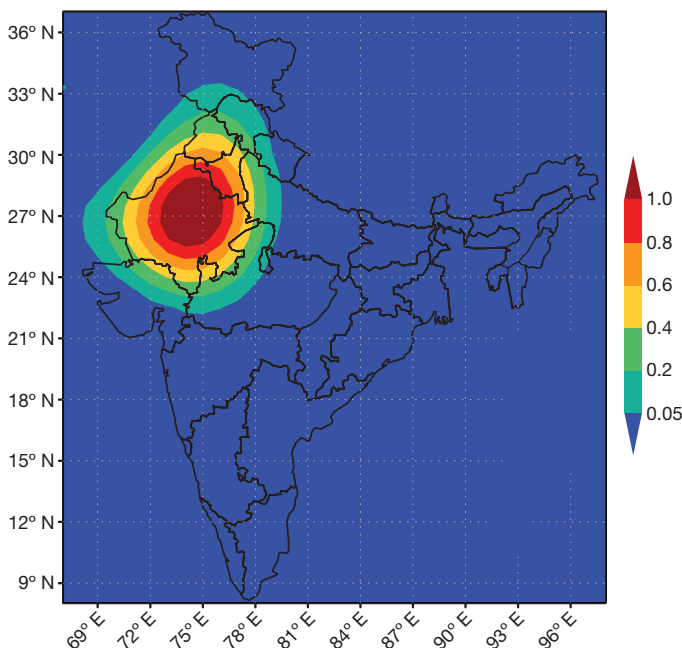
Figure 2 | GRACE averaging function. The unscaled, dimensionless averaging function used to estimate terrestrial water storage changes from GRACE data is mapped.

the terrain shallow; hence, snow is uncommon. However, glaciers abound in the Himalayas, as close as 100 km northeast of Punjab, so we assessed the potential for leakage into our region of the gravity signal associated with non-seasonal glacier melt. We determined that even if 13.4 km[3] yr[−1] of glacier loss (equal to the estimated rate of melt for all of the Himalayas during 1962–2004[19]) occurred along the 150-km stretch of mountains (6% of the Himalayan range) closest to Punjab and Haryana, only 2.8 km[3] yr[−1] (15%) of the estimated TWS trend could be attributed to glacier retreat. We assessed potential leakage from other adjacent trends and included the effects in our error budget.

Surface-water storage also merits consideration. The 203 main reservoirs[20] within or on the border of the three states, plus a large salt lake in Rajasthan, occupy a total area of 4,320 km[2] and have a gross storage capacity of 39.5 km[3]. On the basis of reservoir storage reports from the Indian Central Water Commission, seven of the largest reservoirs, totalling 34.4 km[3] of gross storage capacity, experienced a net increasing trend of 0.5 km[3] yr[−1] during the study period. The total area of rice paddies in the three states is 38,061 km[2] (8.7% of the region), but all are seasonally flooded, shallow-water (<1-m) paddies[21], which dry annually. Hence, if anything, surface-water storage changes probably reduced the rate of TWS depletion observed by GRACE.

Root-zone soil water is often the dominant contributor to TWS variability in temperate regions[22]. To remove its effect and thus isolate groundwater storage changes, we estimated soil-water storage variations by averaging results from five simulations of the Global Land Data Assimilation System (GLDAS)[4]. These simulations used combinations of three land-surface models and three meteorological-forcing (input) data sets. Total soil-column depths in the models ranged from 200 to 340 cm; thus, they did not account for water storage variations in deep unsaturated soil. However, sub-root-zone soil dries only by gravity drainage or by diffusion to drier layers above. The lack of a drying trend in the root zone (as described below) indicates that deep soil-water storage was likewise stable.

Figure 3 shows the resulting 6-yr time series of monthly groundwater storage anomalies as equivalent heights of water averaged over the three-state region. It compares favourably with the TWS and soil-water time series. Its seasonal cycle (Fig. 3, inset) lags soil water by about 40 d and TWS by about 20 d. The mean peak-to-peak amplitude
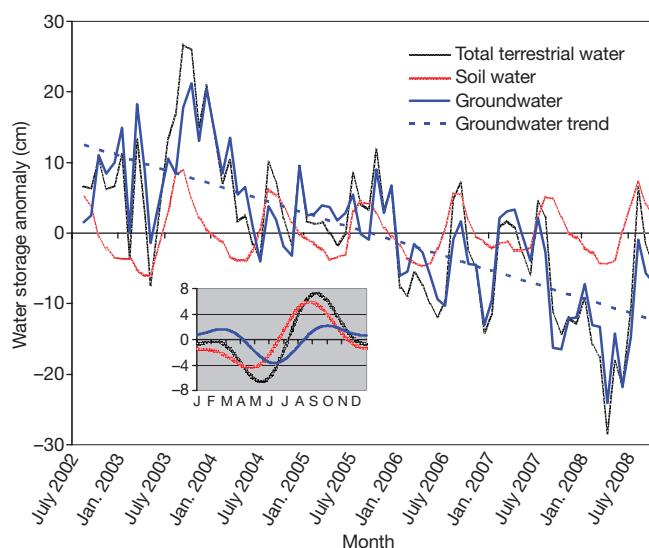
Figure 3 | Monthly time series of water storage anomalies in northwestern India. Monthly time series of anomalies of GRACE-derived total TWS, modelled soil-water storage and estimated groundwater storage, averaged over Rajasthan, Punjab and Haryana, plotted as equivalent heights of water in centimetres. Also shown is the best-fit linear groundwater trend. Inset, mean seasonal cycle of each variable.

is 5.9 cm, compared with 10.3 cm for soil water and 13.8 cm for TWS. These relationships are consistent with those reported in previous studies of soil-water/groundwater covariability[5,14,22]. The soil-water time series reflects rainfall anomalies during the period (discussed below) and exhibits no significant trend. On the other hand, TWS and groundwater decline steadily from 2003 onwards. We calculate the rate of depletion of groundwater to be $4.0 \pm 1.0 \, \text{cm yr}^{-1}$. Assuming[2] a specific yield of 0.12, the regional mean rate of water table decline would be about $0.33 \, \text{m yr}^{-1}$. Local rates of water table decline, which are highly variable, are reported to be as large as $10 \, \text{m yr}^{-1}$ in certain urban areas (Shajan, B., 'NGOs welcome High Court order on water table' (http://www.thehindu.com/2004/01/15/stories/2004011509010400.htm), The Hindu, 15 January 2004). As best can be determined from the coarse GRACE observations, maximum rates of groundwater depletion are centred on Haryana. Groundwater levels also appear to be declining quickly in western Uttar Pradesh, to the east of Haryana. If there is groundwater depletion in Pakistan, to the northwest, it seems to be much less severe.

Although six years is a short period from which to assess a long-term trend with confidence, two pieces of evidence support our conclusion that severe groundwater depletion is occurring as a result of human consumption rather than natural variability. First, the Indian Ministry of Water Resources reports that groundwater withdrawals exceed recharge in the three states we studied[2]. Irrigation accounts for about 95% of the consumption[2]; about 28% of the area is irrigated[23]. Second, there was no shortage of rainfall in the region to cause a natural decline in water storage. For 2002–2008, two observation-based precipitation products[15,16] and an operational atmospheric analysis product[17] indicated a small positive trend in the region, in the range of $2.5–5.1 \, \text{cm yr}^{-1}$. Relative to the climatological mean, 2002 was a dry year, 2003 and 2008 were wet years and rainfall in the other years was within a few centimetres of normal.

GLDAS modelled soil-water fields integrate the effects of precipitation, solar radiation, air temperature and other meteorological factors that directly or indirectly influence groundwater storage[4]. The trend in simulated soil-water storage during the period of study was $0.4 \, \text{cm yr}^{-1}$. This supports the notion that groundwater declines were not caused by natural climate variability. It also confirms that the computed groundwater trend is not a mathematical artefact caused by the subtraction of a large positive soil-water trend from the GRACE-derived TWS trend.

We conclude that withdrawals for irrigation and other uses are depleting the groundwater reserves of Rajasthan, Punjab and Haryana at a rate of $4.0 \pm 1.0 \, \text{cm yr}^{-1}$ equivalent height of water, or $17.7 \pm 4.5 \, \text{km}^3 \, \text{yr}^{-1}$. The Indian Ministry of Water Resources reports that the difference between annual available recharge and annual withdrawals in the region is a $13.2 \, \text{km}^3 \, \text{yr}^{-1}$ deficit[2]. Our result implies that the portion of irrigated water that replenishes the aquifers is less and/or the rate of withdrawal is more than the Indian government has estimated. Apparently, most of the groundwater withdrawn subsequently is lost from the region as a result of increases in run-off and/or evapotranspiration. Between August 2002 and October 2008, the region lost $109 \, \text{km}^3$ of groundwater, which is double the capacity of India's largest reservoir, the Upper Wainganga, and almost triple the capacity of the largest man-made reservoir in the United States, Lake Mead. Depletion is likely to continue until effective measures are taken to curb groundwater demand or until the supply or quality of the resource is diminished to the point at which farmers and residents of the region are forced to react. Severe shortages of potable water, reduced agricultural productivity, conflict and suffering surely would accompany the supply-limited solution.

METHODS SUMMARY

We used 73 monthly GRACE gravity solutions generated by the Center for Space Research at the University of Texas at Austin, consisting of sets of spherical harmonic (Stokes) coefficients, C_{lm} and S_{lm}, to degree l and order m, both

≤ 60. After removing the temporal mean to obtain gravitational anomalies, we filtered[24] each field to reduce noise and converted it to mass in units of equivalent water thickness. We then computed the uncertainty associated with the GRACE measurements[25,26].

To compute the groundwater storage time series, we removed GLDAS[4] estimates of soil-water storage variations from the GRACE TWS[14]. For consistency, we applied degree-60 truncation and the same filtering process to the soil-water time series. Uncertainty in the GLDAS modelled soil-water time series was computed as the standard deviation of results from the five contributing simulations[27].

We then generated time series of monthly TWS, soil-water and groundwater storage in the study region by applying an averaging kernel (Fig. 2) to the Stokes coefficients. The kernel is spatially smoothed primarily as a consequence of three factors: the truncation of the Stokes coefficients to degree 60, the filtering process and convolution with a Gaussian function. When mass variations inside and outside the region differ, such smoothing causes the amplitude of the retrieved signal to be damped[26,28]. To offset this effect, we computed and applied a scaling factor of 1.95. We estimated leakage from the surrounding regions using various scenarios to determine all possible impacts. The total uncertainty estimate for the groundwater trend, $1.0 \, \text{cm yr}^{-1}$, combines errors associated with the GRACE measurements ($0.54 \, \text{cm yr}^{-1}$), GLDAS soil water ($0.38 \, \text{cm yr}^{-1}$), scaling ($0.50 \, \text{cm yr}^{-1}$) and leakage ($0.60 \, \text{cm yr}^{-1}$). An expanded description of the methods used is provided in the online Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Postel, S. in *Water in Crisis: A Guide to the World's Freshwater Resources* (ed. Gleick, P. H.) 56–66 (Oxford Univ. Press, 1993).
2. Central Ground Water Board. *Dynamic Ground Water Resources of India (as on March 2004)* (Ministry of Water Resources, Government of India, 2006).
3. Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F. & Watkins, M. M. GRACE measurements of mass variability in the Earth system. *Science* 305, 503–505 (2004).
4. Rodell, M. *et al.* The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* 85, 381–394 (2004).
5. Changnon, S. A., Huff, F. A. & Hsu, C. F. Relations between precipitation and shallow groundwater in Illinois. *J. Clim.* 1, 1239–1250 (1988).
6. Sturchio, N. C. *et al.* One million year old groundwater in the Sahara revealed by krypton-81 and chlorine-36. *Geophys. Res. Lett.* 31, L05503 (2004).
7. Livingston, M. L. & Garrido, A. Entering the policy debate: an economic evaluation of groundwater policy in flux. *Wat. Resour. Res.* 40, doi:10.1029/2003WR002737 (2004).
8. Kumar, R., Singh, R. D. & Sharma, K. D. Water resources of India. *Curr. Sci.* 89, 794–811 (2005).
9. Briscoe, J. *India's Water Economy: Bracing for a Turbulent Future*. Report No. 34750-IN, viii–xi (World Bank, 2005).
10. Mall, R. K., Gupta, A., Singh, R., Singh, R. S. & Rathore, L. S. Water resources and climate change: an Indian perspective. *Curr. Sci.* 90, 1610–1626 (2006).
11. Zaisheng, H., Hao, W. & Rui, C. *Transboundary Aquifers in Asia with Special Emphasis on China* 10–18 (United Nations Educational, Scientific and Cultural Organization, 2006).
12. Ministry of Water Resources, Government of India. *Central Ground Water Authority* 〈http://cgwb.gov.in/GroundWater/gw_regulation.htm〉 (2009).
13. Yeh, P. J.-F., Swenson, S. C., Famiglietti, J. S. & Rodell, M. Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE). *Wat. Resour. Res.* 42, doi:10.1029/2006WR005374 (2006).
14. Rodell, M. *et al.* Estimating ground water storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol. J.* 15, doi:10.1007/s10040-006-0103-7 (2007).
15. Huffman, G. J. *et al.* The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* 8, 38–55 (2007).
16. Xie, P. P. & Arkin, P. A. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Am. Meteorol. Soc.* 78, 2539–2558 (1997).
17. Persson, A. *User Guide to ECMWF Forecast Products*. Meteorological Bulletin M3.2 (European Centre for Medium-Range Weather Forecasts, 2001).
18. Rodell, M., Chao, B. F., Au, A. Y., Kimball, J. & McDonald, K. Global biomass variation and its geodynamic effects, 1982–1998. *Earth Interact.* 9, 1–19 (2005).
19. Dyurgerov, M. B. & Meier, M. F. *Glaciers and the Changing Earth System: A 2004 Snapshot*. Occasional Paper No. 58 (Institute of Arctic and Alpine Research, University of Colorado, 2005).
20. Central Water Commission, Government of India. *National Register of Large Dams – 2002* (Government of India, 2002).
21. Frolking, S., Yeluripati, J. B. & Douglas, E. New district-level maps of rice cropping in India: a foundation for scientific input into policy. *Field Crops Res.* 98, doi:10.1016/j.fcr.2006.01.004 (2006).

22. Rodell, M. & Famiglietti, J. S. An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE). *Wat. Resour. Res.* **37**, 1327–1340 (2001).
23. Siebert, S. *et al.* Development and validation of the global map of irrigation areas. *Hydrol. Earth Syst. Sci.* **9**, 535–547 (2005).
24. Swenson, S. & Wahr, J. Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.* **33**, doi:10.1029/2005GL025285 (2006).
25. Wahr, J., Swenson, S. & Velicogna, I. The accuracy of GRACE mass estimates. *Geophys. Res. Lett.* **33**, doi:10.1029/2005GL025305 (2006).
26. Swenson, S., &. Wahr, J. Multi-sensor analysis of water storage variations of the Caspian Sea. *Geophys. Res. Lett.* **34**, doi:10.1029/2007GL030733 (2007).
27. Kato, H. *et al.* Sensitivity of land surface simulations to model physics, land characteristics, and forcings, at four CEOP Sites. *J. Meteorol. Soc. Jpn* **87A**, 187–204 (2007).
28. Velicogna, I. & Wahr, J. Acceleration of Greenland ice mass loss in spring 2004. *Nature* **443**, 329–331 (2006).

METHODS

We used 73 monthly GRACE gravity field solutions generated by the Center for Space Research at the University of Texas at Austin[3]. Each monthly gravity field consisted of a set of spherical harmonic (Stokes) coefficients, C_{lm} and S_{lm}, to degree l and order m, both ≤ 60. The C_{20} coefficients showed unreasonable variability, so we replaced them with values derived from satellite laser ranging[29]. After removing the temporal mean to obtain gravitational anomalies, we filtered[24] each monthly field to reduce noise ('striping') and converted it to mass in units of equivalent water thickness.

We corrected the GRACE mass anomalies for the solid-Earth contributions generated by the high-latitude Pleistocene deglaciation using an independent model[30]. However, the contribution was negligible in this analysis.

We used GLDAS[4] time series of soil-water storage to isolate groundwater storage variations from the GRACE TWS anomalies[14]. For consistency, before we calculated Stokes coefficients for the soil water storage we truncated them to degree 60 and applied the same filtering process as used for the GRACE fields.

We used the filtered, monthly coefficients to generate time series for TWS, soil water and groundwater storage in the study region. To do this we constructed an averaging kernel by convolving a Gaussian function with a gridded map of the combined area of Punjab, Haryana (including Delhi) and Rajasthan (438,296 km²). After testing various radii for the Gaussian function, from 300 km to 0 km, we selected an exact (0-km) averaging function (that is, no Gaussian averaging) because it minimized uncertainty and was generally best suited for our region (see below).

We applied the kernel to the Stokes coefficients to estimate mass changes averaged over the region. Ideally, the kernel should take the value 1 within a study region and 0 outside. In practice, the kernel is smoothed as a consequence of three factors: the truncation to degree 60 (the maximum degree for the GRACE Stokes coefficients), the filtering process and the convolution with the Gaussian function. Such smoothing causes an amplitude damping of the recovered mass. To obtain a realistic estimate of the mass changes, the damping of the signal amplitude caused by smoothing must be mitigated[26,28]. Also, we must account for potential leakage of mass changes from adjacent regions. In particular, the leakage effects may or may not be significant. In cases where mass changes are uniform across the region's boundaries, bias is minimal[25,26] and the signal recovered using the averaging kernel does not need to be rescaled because the smoothing effects are compensated by leakage of identical variations from the outside regions. However, if mass variability immediately outside the region is uncorrelated with that inside or they are negatively correlated or have different amplitudes, the recovered signal will be biased[28]. This must be evaluated on a case-by-case basis. We determined, on the basis of the GLDAS soil-water fields and the GRACE fields, that seasonal variations in TWS inside and outside the region are well correlated. In fact, the exact average of the GLDAS soil water over the study region and the average calculated when applying the truncation, filtering and Gaussian averaging differed only by few per cent.

On the other hand, the secular change (trend) varies between the inside and the outside of the study region, with the negative anomaly mainly concentrated inside the region. Therefore, to recover the total mass trend we rescaled the time series to account for amplitude damping[26,28]. The scaling factor was estimated by distributing a synthetic mass-change signal uniformly over the study region, processing it in the same manner as the GRACE data (that is, converting it to the spectral domain, truncating it to degree 60 and filtering and spatially averaging it) and comparing the retrieved signal with the original synthetic signal. We then multiplied the GRACE time series trend by this factor, which was determined to be 1.95. The scaling depends on how mass change is distributed within the region; therefore, to account for the uncertainty associated with our representation of that distribution, we calculated the sensitivity to different mass-change distributions within the region. We found that mass distribution effects caused an uncertainty of 0.5 cm yr^{-1}, which we included in our final error budget.

We also evaluated the leakage from mass-change signals in the surrounding regions. To estimate the leakage from mass loss associated with non-seasonal Himalayan glacier melt, we distributed a trend of 13.4 km³ yr^{-1}, equal to the estimated rate of melt for the all of the Himalayas during 1962–2004[19], along the 150-km stretch of glaciers closest to Punjab and Haryana, and applied our averaging function and scaling. This is a highly conservative approach, given that 150 km is only 6% of the length of the mountain range. The portion of that signal that leaked into our region was 2.8 km³ yr^{-1}, or about 15% of the estimated groundwater trend. Because that is almost surely a significant overestimate, we did not use it to adjust the estimated groundwater trend, but continued with a more careful analysis of leakage effects. In that analysis, we included leakage from all TWS trends in the surrounding region that were detected by GRACE. We investigated different leakage scenarios to determine all possible effects on our final estimate. In the most likely scenario, the leakage effects nearly cancel. Nevertheless, uncertainty due to leakage was estimated to be 0.6 cm yr^{-1}, which is included in the final error budget.

The decision to use an exact (radius, $R = 0$ km) Gaussian function was based on the following reasoning. We first evaluated the effects of filtering in the studied region by comparing trend maps obtained by applying 300-, 250-, 200-, 150-, 100-, 50- and 0-km Gaussian averaging. Even if $R = 0$ km and there is no Gaussian smoothing, the resulting map will still be smoothed as a consequence of the necessary truncation of the GRACE solutions to degree 60. We found that the seven smoothing radii produced insignificant differences in our region; thus, none of them were eliminated from consideration.

We constructed seven different averaging kernels by convolving Gaussian functions for each of the seven radii with a gridded map of the region of interest. We calculated TWS time series using the seven kernels, and we estimated the corresponding scaling factors and measurement errors. Both the scaling factor and the error are a function of the specific kernel (note that in the case of $R = 0$ km, use of the scaling factor is necessitated only by the truncation to degree 60 and filtering of the GRACE solutions). For example, the averaging kernel produced by convolving a Gaussian function with $R = 250$ km with a gridded map will deliver results that are more smoothed than in the case in which $R = 200$ km (that is, the associated scaling factor will be larger for $R = 250$ km than for $R = 200$ km) but will have smaller measurements errors.

A priori, we did not know which would be the best averaging function because of this trade-off between smoothing and errors. Also, it is possible for the net effect of striping to nearly cancel itself out, depending on the specific location. We compared the TWS time series, scaling factor and errors for each of the seven averaging functions and found that the GRACE measurement errors (defined by the scatter of the monthly values about their seasonal cycle) were not significantly larger in the case of $R = 0$ km than in the case of any of the greater radii (again, truncation to degree 60 produces some smoothing, even in the case of $R = 0$ km).

Uncertainties in the Stokes coefficients were determined by assuming that the scatter of the monthly values about their seasonal cycle is due entirely to errors[25]. This represents the upper bound on the random component of the error. The estimate is conservative, because intra-annual variations in the signal will be interpreted as error. The 1σ error estimates in the spatially averaged GRACE time series were then calculated from the uncertainty in the individual Stokes coefficients[26]. The scaled root-mean-squared error in the trend is 0.54 cm yr^{-1}.

Uncertainty in the GLDAS modelled soil-water time series was computed as the standard deviation of results from the five contributing simulations[27]. This calculation was done separately for the seasonal and secular components. The uncertainty in each monthly soil-water estimate for the study region is 1.71 cm and uncertainty in the trend is 0.38 cm yr^{-1}. The total error estimate for the ground water trend, 1.0 cm yr^{-1}, combines the GRACE measurement errors, the GLDAS error, the scaling error and the leakage error, under the assumption that the individual errors are uncorrelated.

29. Cheng, M. K. & Tapley, B. D. Variations in the Earth's oblateness during the past 28 years. *J. Geophys. Res.* **109**, B09402 (2004).

30. Paulson, A., Zhong, S. & Wahr, J. Inference of mantle viscosity from GRACE and relative sea level data. *Geophys. J. Int.* **171**, 497–508 (2006).

LETTERS

Global electromagnetic induction constraints on transition-zone water content variations

Anna Kelbert[1], Adam Schultz[1] & Gary Egbert[1]

Small amounts of water can significantly affect the physical properties of mantle materials, including lowering of the solidus[1], and reducing effective viscosity[2] and seismic velocity[3]. The amount and distribution of water within the mantle thus has profound implications for the dynamics and geochemical evolution of the Earth[4,5]. Electrical conductivity is also highly sensitive to the presence of hydrogen in mantle minerals[6]. The mantle transition zone minerals wadsleyite and ringwoodite in particular have high water solubility[4], and recent high pressure experiments show that the electrical conductivity of these minerals is very sensitive to water content[7–9]. Thus estimates of the electrical conductivity of the mantle transition zone derived from electromagnetic induction studies have the potential to constrain the water content of this region. Here we invert long period geomagnetic response functions to derive a global-scale three-dimensional model of electrical conductivity variations in the Earth's mantle, revealing variations in the electrical conductivity of the transition zone of approximately one order of magnitude. Conductivities are high in cold, seismically fast, areas where slabs have subducted into or through the transition zone. Significant variations in water content throughout the transition zone provide a plausible explanation for the observed patterns. Our results support the view[10,11] that at least some of the water in the transition zone has been carried into that region by cold subducting slabs.

Electromagnetic induction sounding of the Earth's mantle entails inverting impedance functions calculated from the ratios of magnetic (geomagnetic depth sounding) or electric and magnetic (magnetotelluric) field components. Magnetic variations with periods of approximately 2–100 days, which are associated with the relaxation phase of disturbed storm time (D_{st}) variations of the magnetospheric equatorial ring current, provide good resolution in and just below the mantle transition zone[12]. Until recently, the principal constraints on electrical conductivity at these depths have derived from one-dimensional inversion of geomagnetic depth sounding and/or magnetotelluric responses (Methods). Several recent studies have discussed significant spatial variability in transition-zone electrical conductivity, and its implications for water content[13,14], but these have all been based on one-dimensional local interpretations. Regional magnetotelluric surveys, which are generally interpreted with two-dimensional models, typically extend to periods of no longer than 10^4 s (10^5 s for seafloor magnetotellurics), thus essentially limiting studies to the upper mantle and top of the transition zone[15]. Initial efforts at three-dimensional inversion of very long period geomagnetic depth sounding and magnetotelluric data, including a regional model of the North Pacific[16,17] and a global-scale inversion using an approximate forward modelling scheme[18], have so far provided very limited views of whole mantle conductivity.

To obtain a global three-dimensional model of mantle electrical conductivity, we use a frequency-domain, regularized nonlinear conjugate gradient inversion for the spherical Earth[12]. We parameterize the electrical conductivity in the mantle (depths of 12.65–1,600 km) by a series of spherical shells, in each of which lateral conductivity variations about a one-dimensional prior model[19] are represented by a spherical harmonic expansion of degree and order nine. The prior model has been discretized into 12 layers[12], allowing for conductivity jumps at 410, 520 and 670 km depths, related to the major mineral phase transitions. The lower four layers (1,600 km and deeper) were constrained to be homogeneous. The inversions were regularized by minimizing deviations from the prior model, with a norm that penalized higher degree and order terms, and jumps between layers. For our study, we have used a compilation of 59 observatories equator-ward of $60°$ geomagnetic latitude from global observatory data sets[20,21]. Responses at 28 periods from 5 to 106.7 days were inverted (Methods Summary).

An extensive set of computational experiments were performed, using different combinations of vertical and horizontal regularizations, damping parameters, and degrees and orders of spherical harmonic model parameterization, with the effect of isolated or possibly anomalous observatories tested by omitting these from the inversion. The large-scale pattern of mid-mantle heterogeneities that emerges from all of these tests is robust, although details of the locations of boundaries between regions of differing conductivity, as well as the amplitudes of the heterogeneities, vary somewhat. We have also explored the trade-off between data fit and model norm. The full data set can be fitted to a normalized root mean square (r.m.s.) of 1.06, only slightly larger than the idealized target normalized r.m.s. of 1. Bearing in mind that modelling inaccuracies and source complications are not accounted for in the statistical data error estimates, we choose a slightly poorer fitting (r.m.s. = 1.14) model for further discussion (Fig. 1). Better fitting models exhibit similar spatial structure, but with larger amplitude anomalies.

The pattern of lateral heterogeneities in our models appears continuous through and below the mantle transition zone. This may result from the limited vertical resolution of this data set, and the regularization employed. The induction data cannot distinguish between a thin layer of very high conductivity at the base of the transition zone and more moderate conductivities spreading into the lower mantle. Inversion of synthetic data with an identical distribution of sites and comparable signal and noise characteristics demonstrates[12] that in regions with good data coverage, large-scale structures at depths of ~500–1,600 km are resolved relatively well, even if shorter wavelength conductivity variations are present. However, the deeper structures should be interpreted with some caution owing to possible distortion or masking by the high electrical conductivities of the transition zone. Here we focus on the structure in, and just below, the transition zone (depths <900 km). In regions with poor data coverage, including most of Africa, South America, and the Indian and South Pacific Oceans, conductivity is poorly constrained at all mantle depths[12].

[1]College of Oceanic & Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331-5503, USA.
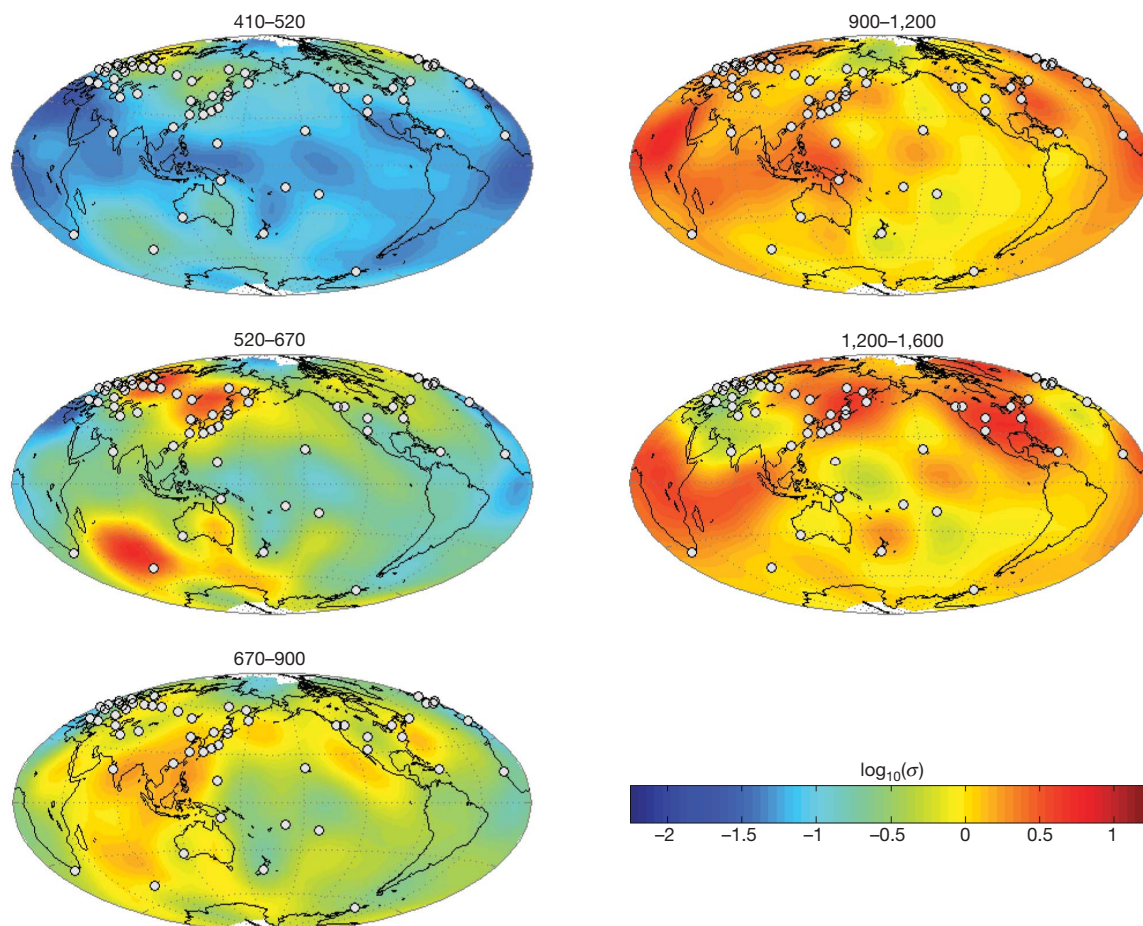
Figure 1 | Regularized degree and order nine electrical conductivity inverse solution. The solution has been obtained by applying a global inversion algorithm[12] to C responses from the mid-latitude subset of the global observatory data set[20], supplemented by six C responses from ref. 21, with correction for near-surface effects[30]. Here, σ stands for electrical conductivity. The 59 data locations are denoted by white dots; the depths from the Earth's surface are indicated in km. The normalized r.m.s. misfit of this model is 1.14. Regions of poor spatial resolving power include most of Africa and the Indian Ocean, the Southern Pacific and South America.

The most notable feature in Fig. 1 at transition-zone depths is a band of enhanced conductivity along the circum-Pacific margin, extending from western North America, through the Aleutian arc and East Asia, and into the Indian Ocean and Australia. Most notably, high conductivity is registered beneath Japan, eastern China, and in the areas of the Izu-Bonin and Tonga slabs. The central Pacific basin is more than an order of magnitude more resistive than the surrounding conductive regions. Lack of continuity of this conductive feature, and its absence from the remaining parts of the circum-Pacific margin (for example, beneath South America) may be the result of the significant gaps in data coverage. The high conductivity areas in and below the mantle transition zone correlate well with the present-day distribution of cold, and seismically fast, subducted slabs[22], at least in areas with good model resolution. We also note positive correlation with maps of transition-zone thickness, and depth to the 410 km discontinuity[23].

These correlations strongly suggest that the high conductivities at transition-zone depths are linked to subducted oceanic lithosphere. Based on temperature alone, this cold subducted material should be relatively resistive[7,8]. Water, carried into the transition zone with the subducting slabs, provides a plausible explanation for the elevated conductivities. Transition-zone minerals have high water solubility[4] (about 2 wt%), and 1 wt% water raises the conductivity of dry wadsleyite and ringwoodite by more than an order of magnitude[7–9]. Water may also elevate conductivity indirectly by inducing melting in or immediately above or below the transition zone[5,14]. Oxidized carbon may also have a role in enhancing mantle conductivity (by

inducing partial melt or otherwise), although evidence to date suggests[24] that such enhancements would be restricted to shallow transition-zone depths or above (Methods). Elevated iron content[25] would also increase conductivity of these transition-zone minerals.

The water content of the mantle transition zone may be roughly estimated using high-pressure laboratory measurements of electrical conductivity in samples of the transition-zone minerals of varying hydrogen content[7–9]. In Fig. 2, we compare electrical conductivity profiles from our three-dimensional inverse solution to those derived from the relation between conductivity and water content given in ref. 8, with water content varying from 0 to 1 wt%. Assuming the results of ref. 8, 0.5–1 wt% water might be required to achieve the highest conductivities in the lower transition zone (0.5–2.0 S m^{-1}, for example, beneath northeastern China and Japan), while the more resistive areas are consistent with dry mantle conditions, with no more than 0.1 wt% water. Slightly lower water content would be inferred in the conductive anomalies from the results of ref. 9. More rigorous quantification of water content must await resolution of uncertainties in laboratory measurements, and would also require incorporating additional geophysical constraints and allowing for other possible contributing factors, such as partial melt, as well as for lateral temperature variations.

The significant spatial variations in the transition zone, with the highest conductivities in areas with significant recent subduction input, suggest that water is carried into the transition zone with the hydrated parts of the lithosphere. This possibility has been previously suggested by a number of studies[10,26]. Water transport into the deep
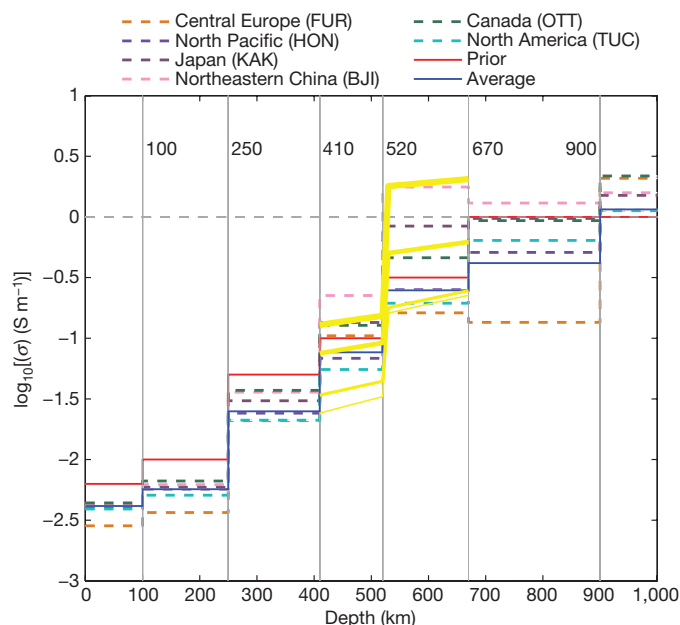
Figure 2 | Global and regional electrical conductivity profiles, based on the three-dimensional inverse solution presented in Fig. 1. The dashed lines correspond to the profiles beneath a set of locations, representative of geographical regions. The blue solid line is the global average. The red solid line represents the prior one-dimensional model used for the inversion[19]. The four yellow lines indicate the existing mineral physics constraints[8] as a function of water content (from the bottom up, in wt%: 0; 0.1, 0.5, 1.0). The three-letter abbreviations refer to the INTERMAGNET geomagnetic observatory codes[27].

mantle is expected to be particularly effective for the cold Pacific plate subducting along the Japan trench, where hydrous phases may remain stable to depths greater than 300 km (ref. 11). Notably this region is associated with particularly high conductivity in our models. The presence of extensive areas of low conductivity, which must be dry, suggests that mechanisms transporting water out of the transition zone must also be active. Further research will be required to clarify the complete transition-zone water cycle.

The model presented here is to our knowledge the first global three-dimensional electrical conductivity model capable of providing constraints on the lateral variations in mantle water content. However, the data set used in this study provides limited vertical and horizontal spatial resolution. The former could be improved by incorporating data from new high quality geomagnetic observatories (especially in the Southern Hemisphere and in Asia) which have now been occupied long enough to provide reliable response functions[27]. Data from satellites such as Champ and Ørsted (and the upcoming SWARM mission) will improve spatial coverage further[28]. Vertical resolution will always be limited by the diffusive nature of electromagnetic fields in a conductor, but could be considerably improved in the upper mantle (including the top of the transition zone) by extending the data set to shorter periods[12]. Interpreting such data will require more sophisticated models for magnetospheric and ionospheric source currents, which are spatially more complex at these periods. More realistic source models would allow more complete use of the available geomagnetic data, including fitting anomalies in all field components[20], and interpreting data from higher geomagnetic latitudes. These extensions will no doubt allow the model presented here to be refined significantly.

METHODS SUMMARY

For our analysis, we use a regularized nonlinear conjugate gradient inversion[12] for global electromagnetic induction studies, based on an adjoint formulation of a staggered-grid finite difference frequency domain numerical solution of Maxwell's equations for spherical geometry[29].

Global observatory response functions from ref. 20 (known as C responses[21]) are inverted for conductivity under the assumption that external source fields can be approximated as a zonal dipole[21]. Currents in the auroral ionosphere that are coherent with the magnetospheric equatorial ring current responsible for the dipole source cause this approximation to degrade[20] at geomagnetic latitudes pole-ward of ± 45–$50°$. To minimize artefacts due to these source field complications, responses were corrected for the effect of auroral ionosphere current systems[20], and the inversion was restricted to the 53 observatories equator-ward of $60°$ geomagnetic latitude. The associated jackknife error estimates have median amplitudes of roughly 10%, with larger errors at higher latitudes. The data set is augmented by C responses from ref. 21 at six mid-latitude locations to fill in the under-sampled geographical areas.

Coastal observatory responses can be significantly distorted at periods of up to 20 days by the large lateral conductivity contrasts associated with the oceans and underlying sediments. Accurate modelling of these effects requires high numerical resolution (grids of the order of $1° \times 1°$)[30], whereas practicality dictates a much coarser ($10° \times 10°$) grid for the inversions discussed here. To address this issue, we corrected the C responses as suggested in ref. 30, using a high resolution ($1° \times 1°$) grid to model near-surface conductive heterogeneities. Tests showed that although fits to some coastal observatories were significantly improved with these corrections, there was little impact on large-scale structure of the resulting conductivity models. This result, and further tests with synthetic data, gives us confidence that even with our approximate treatment this complication does not unduly affect our results.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144,** 93–108 (1996).
2. Dixon, J. E., Dixon, T. H., Bell, D. R. & Malservisi, R. Lateral variation in upper mantle viscosity: role of water. *Earth Planet. Sci. Lett.* **222,** 451–467 (2004).
3. Karato, S.-I. & Jung, H. Water, partial melting and the origin of the seismic low velocity and high attenuation zone in the upper mantle. *Earth Planet. Sci. Lett.* **157,** 193–207 (1998).
4. Williams, Q. & Hemley, R. J. Hydrogen in the deep Earth. *Annu. Rev. Earth Planet. Sci.* **29,** 365–418 (2001).
5. Bercovici, D. & Karato, S.-i. Whole-mantle convection and the transition-zone water filter. *Nature* **425,** 39–44 (2003).
6. Karato, S. The role of hydrogen in the electrical conductivity of the upper mantle. *Nature* **347,** 272–273 (1990).
7. Huang, X., Xu, Y. & Karato, S.-i. Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite. *Nature* **434,** 746–749 (2005).
8. Yoshino, T., Manthilake, G., Matsuzaki, T. & Katsura, T. Dry mantle transition zone inferred from the conductivity of wadsleyite and ringwoodite. *Nature* **451,** 326–329 (2008).
9. Karato, S. & Dai, L. Comments on "Electrical conductivity of wadsleyite as a function of temperature and water content" by Manthilake *et al. Phys. Earth Planet. Inter.* (in the press).
10. Ohtani, E., Litasov, K., Hosoya, T., Kubo, T. & Kondo, T. Water transport into the deep mantle and formation of a hydrous transition zone. *Phys. Earth Planet. Inter.* **143,** 255–269 (2004).
11. Kawakatsu, H. & Watada, S. Seismic evidence for deep-water transportation in the mantle. *Science* **316,** 1468–1471 (2007).
12. Kelbert, A., Egbert, G. D. & Schultz, A. Non-linear conjugate gradient inversion for global EM induction: resolution studies. *Geophys. J. Int.* **173,** 365–381 (2008).
13. Hae, R., Ohtani, E., Kubo, T., Koyama, T. & Utada, H. Hydrogen diffusivity in wadsleyite and water distribution in the mantle transition zone. *Earth Planet. Sci. Lett.* **243,** 141–148 (2006).
14. Toffelmier, D. A. & Tyburczy, J. A. Electromagnetic detection of a 410-km-deep melt layer in the southwestern United States. *Nature* **447,** 991–994 (2007).
15. Booker, J. R., Favetto, A. & Pomposiello, M. C. Low electrical resistivity associated with plunging of the Nazca flat slab beneath Argentina. *Nature* **429,** 399–403 (2004).
16. Fukao, Y., Koyama, T., Obayashi, M. & Utada, H. Trans-Pacific temperature field in the mantle transition region derived from seismic and electromagnetic tomography. *Earth Planet. Sci. Lett.* **217,** 425–434 (2004).
17. Koyama, T. *et al. Water Content in the Mantle Transition Zone Beneath the North Pacific Derived from the Electrical Conductivity Anomaly* 171–179 (AGU Geophys. Monograph Series 168, 2006).
18. Schultz, A. & Pritchard, G. in *Three Dimensional Electromagnetics* (eds Spies, B. & Oristaglio, V.) 451–476 (Society of Exploration Geophysicists, 1999).
19. Kuvshinov, A. & Olsen, N. A global model of mantle conductivity derived from 5 years of CHAMP, Ørsted, and SAC-C magnetic data. *Geophys. Res. Lett.* **33,** 18301, doi:10.1029/2006GL027083 (2006).
20. Fujii, I. & Schultz, A. The 3D electromagnetic response of the Earth to ring current and auroral oval excitation. *Geophys. J. Int.* **151,** 689–709 (2002).

21. Schultz, A. & Larsen, J. C. On the electrical conductivity of the mid-mantle. 1. Calculation of equivalent scalar magnetotelluric response functions. *Geophys. J. Int.* **88,** 733–761 (1987).
22. Fukao, Y., Widiyantoro, S. & Obayashi, M. Stagnant slabs in the upper and lower mantle transition region. *Rev. Geophys.* **39,** 291–324 (2001).
23. Lawrence, J. F. & Shearer, P. M. Imaging mantle transition zone thickness with *SdS-SS* finite-frequency sensitivity kernels. *Geophys. J. Int.* **174,** 143–158 (2008).
24. Dasgupta, R., Hirschmann, M. M. & Withers, A. C. Deep global cycling of carbon constrained by the solidus of anhydrous, carbonated eclogite under upper mantle conditions. *Earth Planet. Sci. Lett.* **227,** 73–85 (2004).
25. Yoshino, T. & Katsura, T. Effect of iron content on electrical conductivity of ringwoodite, with implications for electrical structure in the transition zone. *Phys. Earth Planet. Inter.* (in the press).
26. Maruyama, S. & Okamoto, K. Water transportation from the subducting slab into the mantle transition zone. *Gondwana Res.* **11,** 148–165 (2007).
27. INTERMAGNET. Participating observatories (map). 〈http://www.intermagnet.org/Imomap_e.html〉.
28. Olsen, N. *et al.* CHAOS — a model of the Earth's magnetic field derived from CHAMP, Ørsted, and SAC-C magnetic satellite data. *Geophys. J. Int.* **166,** 67–75 (2006).
29. Uyeshima, M. & Schultz, A. Geomagnetic induction in a heterogeneous sphere: a new three-dimensional forward solver using a conservative staggered-grid finite difference method. *Geophys. J. Int.* **140,** 636–650 (2000).
30. Kuvshinov, A. V., Olsen, N., Avdeev, D. B. & Pankratov, O. V. Electromagnetic induction in the oceans and the anomalous behaviour of coastal C-responses for periods up to 20 days. *Geophys. Res. Lett.* **29,** 1595, doi:10.1029/2001GL014409 (2002).

METHODS

Our model exhibits a general agreement with results from previous deep mantle electromagnetic induction studies. At shallow depths where our data has little sensitivity, that is, in the upper 200 km of the Pacific mantle, there is persuasive evidence for enhanced conductivity due to dissolved hydrogen[31]. The conductivities beneath the central North Pacific have been shown to rise from $0.01\,\mathrm{S\,m^{-1}}$ at a depth of 200 km to just over $0.1\,\mathrm{S\,m^{-1}}$ at the bottom of the transition zone, then to $1\,\mathrm{S\,m^{-1}}$ at about 1,000 km depth[32,33]. These results are consistent with our model at depths $\sim$500 km and below, resolved[12] with the period range of our data (Figs 1, 2). Similar results have been obtained beneath the Canadian Shield[34]. However, conductivities in the transition zone beneath tectonic North America[35] are about half an order of magnitude greater than beneath the Canadian Shield. This location is right at the border between a conductive and a resistive region in our model. Northeastern China has been shown[36] to be at least one order of magnitude more conductive than the central North Pacific, in agreement with our 3D results (Fig. 2). Other profiles[37,38] represent the average between the central North Pacific and the higher conductivity areas resolved by our model (Fig. 1), such as tectonic western North America, northeastern China and Japan. The variability beneath the North Pacific has been explicitly discussed before. For example, Hae *et al.*[13] have estimated that water content variability of 0.03–0.4 wt% would explain electrical conductivity variations in the upper transition zone (400–550 km depth) beneath the Philippine Sea. Although our results for central Europe are consistent with published 1D profiles[39,40], our model shows a steep gradient in this area, with conductivity increasing to the northeast. This feature, which coincides with the Trans European suture zone, has been noted before[41] and may result from lower mantle upwelling[42]. However, further study is required to rule out the possibility of external source contamination at the high geomagnetic latitudes of Northern Europe.

Finally, recent studies of the conductivity of molten carbonates (carbonatites)[43], depleted in Li and rich in Ca, indicate an enhancement in conductivity of three orders of magnitude relative to molten silicates under similar upper mantle conditions. Both geophysical imaging and petrological sampling are consistent with the presence of such highly conducting carbonatites at depths up to $\sim$300 km beneath the East Pacific Rise[43]. However, current evidence suggests carbonates would be unstable at depths relevant to this study, greater than $\sim$400 km (refs 24, 44).

31. Evans, R. L. *et al.* Geophysical evidence from the MELT area for compositional controls on oceanic plates. *Nature* **437,** 249–252 (2005).
32. Lizarralde, D., Chave, A., Hirth, G. & Schultz, A. Northeastern Pacific mantle conductivity profile from long-period magnetotelluric sounding using Hawaii-to-California submarine cable data. *J. Geophys. Res.* **100** (B9), 17837–17854 (1995).
33. Neal, S. L., Mackie, R. L., Larsen, J. C. & Schultz, A. Variations in the electrical conductivity of the upper mantle beneath North America and the Pacific Ocean. *J. Geophys. Res.* **105** (B4), 8229–8242 (2000).
34. Schultz, A., Kurtz, R. D., Chave, A. D. & Jones, A. G. Conductivity discontinuities in the upper-mantle beneath a stable craton. *Geophys. Res. Lett.* **20,** 2941–2944 (1993).
35. Egbert, G. D. & Booker, J. R. Very long period magnetotellurics at Tucson observatory — implications for mantle conductivity. *J. Geophys. Res.* **97,** 15099–15112 (1992).
36. Ichiki, M. *et al.* Upper mantle conductivity structure of the back-arc region beneath northeastern China. *Geophys. Res. Lett.* **28,** 3773–3776 (2001).
37. Utada, H., Koyama, T., Shimizu, H. & Chave, A. D. A. Semi-global reference model for electrical conductivity in the mid-mantle beneath the north Pacific region. *Geophys. Res. Lett.* **30,** 1194–1197 (2003).
38. Kuvshinov, A., Utada, H., Avdeev, D. & Koyama, T. 3-D modelling and analysis of D_{st} C-responses in the North Pacific Ocean region, revisited. *Geophys. J. Int.* **160,** 505–526 (2005).
39. Olsen, N. The electrical conductivity of the mantle beneath Europe derived from C-responses from 3 to 720 hr. *Geophys. J. Int.* **133,** 298–308 (1998).
40. Tarits, P., Hautot, S. & Perrier, F. Water in the mantle: results from electrical conductivity beneath the French Alps. *Geophys. Res. Lett.* **31,** 6612, doi:10.1029/2003GL019277 (2004).
41. Semenov, V. Y. & Jozwiak, W. Lateral variations of the mid-mantle conductance beneath Europe. *Tectonophysics* **416,** 279–288 (2006).
42. Goes, S., Spakman, W. & Bijwaard, H. A lower mantle source for central European volcanism. *Science* **286,** 1928–1931 (1999).
43. Gaillard, F., Malki, M., Iacono-Marziano, G., Pichavant, M. & Scaillet, B. Carbonatite melts and electrical conductivity in the asthenosphere. *Science* **322,** 1363–1365 (2008).
44. Hammouda, T. High-pressure melting of carbonated eclogite and experimental constraints on carbon recycling and storage in the mantle. *Earth Planet. Sci. Lett.* **214,** 357–368 (2003).

LETTERS

Phase-locking and environmental fluctuations generate synchrony in a predator–prey community

David A. Vasseur[1] & Jeremy W. Fox[2]

Spatially synchronized fluctuations in system state are common in physical and biological systems ranging from individual atoms[1] to species as diverse as viruses, insects and mammals[2–10]. Although the causal factors are well known for many synchronized phenomena, several processes concurrently have an impact on spatial synchrony of species, making their separate effects and interactions difficult to quantify. Here we develop a general stochastic model of predator–prey spatial dynamics to predict the outcome of a laboratory microcosm experiment testing for interactions among all known synchronizing factors: (1) dispersal of individuals between populations; (2) spatially synchronous fluctuations in exogenous environmental factors (the Moran effect); and (3) interactions with other species (for example, predators) that are themselves spatially synchronized. The Moran effect synchronized populations of the ciliate protist *Tetrahymena pyriformis*; however, dispersal only synchronized prey populations in the presence of the predator *Euplotes patella*. Both model and data indicate that synchrony depends on cyclic dynamics generated by the predator. Dispersal, but not the Moran effect, 'phase-locks' cycles, which otherwise become 'decoherent' and drift out of phase. In the absence of cycles, phase-locking is not possible and the synchronizing effect of dispersal is negligible. Interspecific interactions determine population synchrony, not by providing an additional source of synchronized fluctuations, but by altering population dynamics and thereby enhancing the action of dispersal. Our results are robust to wide variation in model parameters representative of many natural predator–prey or host–pathogen systems. This explains why cyclic systems provide many of the most dramatic examples of spatial synchrony in nature.

Understanding what causes populations to fluctuate in synchrony is important because synchrony can increase extinction risk in metapopulations[11,12], facilitate recolonization waves in cyclic populations[13], and increase or decrease the stability of food webs[14,15]. Although the synchronizing effects of dispersal and the Moran effect have been well studied[16–20], relatively little effort has focused on the operation of these mechanisms in the presence of species interactions (but see refs 13, 20–23). Besides providing an additional source of synchronized fluctuations, interspecific interactions can introduce high-amplitude cycles in population dynamics through nonlinear feedbacks and time delays[24]. Many of the most dramatic examples of spatial synchrony in nature come from species or systems that cycle through time, including measles and pertussis outbreaks, larch bud moth outbreaks, Fennoscandian voles and Canadian lynx–hare cycles[2,3,7–10], suggesting that interspecific interactions generating cycles may be a crucial ingredient to understanding spatial synchrony.

We predicted the separate and interactive effects of different processes on spatial synchrony by developing a stochastic spatial predator–prey model based on the standard Rosenzweig–MacArthur model

(Methods). Our model incorporated key features of predator–prey biology (density-dependent prey growth and time-limited predator activity budget) known to drive the dynamics of many natural predator–prey systems[25]. We simulated a factorial experiment crossing the presence/absence of dispersal, the Moran effect and the predator to examine their effects on prey synchrony. We conducted 10,000 Monte Carlo simulations of this experiment by varying five key model parameters randomly so as to thoroughly sample a large region of the ecologically relevant parameter space (see Supplementary Information).

The model predicts that the Moran effect synchronizes prey dynamics, but dispersal significantly increases prey synchrony only with predators present (Fig. 1a, b and Supplementary Information). The dispersal–predator interaction arises because predators generate oscillations that are easily synchronized by dispersal. The dispersal–predator interaction is a statistical signature of what is commonly described in dynamical systems theory as 'phase-locking', whereby coupling of periodic or chaotic oscillators entrains their phases and periods[1–4,26,27]. In the absence of dispersal, initially synchronous cycles desynchronize ('decohere') when predator and prey simultaneously reach low densities; subtle differences in noise at the nadir of the cycle can lead to large differences in the timing of subsequent peaks and consequently loss of synchrony (Fig. 2b and Supplementary Information). In the presence of dispersal, cycles tend to emerge from the nadir in lock-step because any differences in density are eliminated by a net flow of individuals from the higher-density to the
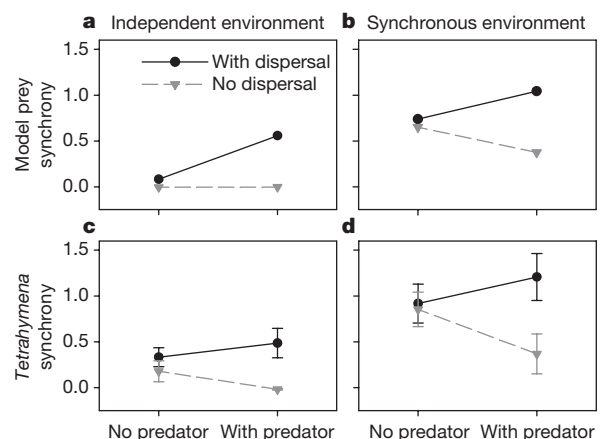
Figure 1 | Three-way interaction plot of the impact of dispersal, the Moran effect and predators on *Tetrahymena* synchrony. a–d, Results from our theoretical model (**a, b**) and our experimental microcosms (**c, d**) are shown. Points represent the mean *z*-transformed Pearson cross-correlation (±1 s.e.m.). In the experiment, both dispersal and the Moran effect significantly increase synchrony, but dispersal does so only in the presence of predators.

[1]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520, USA. [2]Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada.
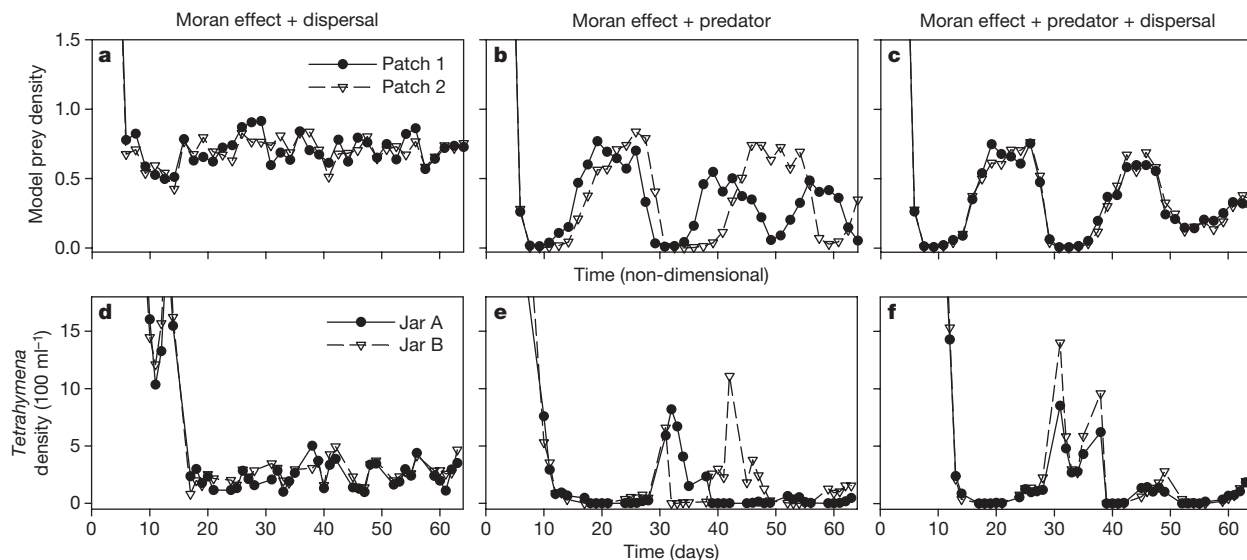
Figure 2 | Temporal dynamics of *Tetrahymena*. a–f, Results from pairs of jars experiencing the Moran effect and dispersal (**a, d**), predators (**b, e**), or predators and dispersal (**c, f**) are shown. Panels **a–c** give illustrative dynamics from our theoretical model, and **d–f** give illustrative dynamics from our experimental microcosms. Adding predators to the theoretical model and experimental *Tetrahymena* cultures generates cyclic dynamics. In the presence of dispersal, cyclic dynamics are entrained by phase-locking. Across panels **a–c** the same initial conditions and stochastic perturbations were used to highlight the effect of predators and dispersal.

lower-density patch (Fig. 2c and Supplementary Information); the greater the density difference between patches the stronger the flow. In contrast, when predators are absent, prey densities exhibit stochastic fluctuations around carrying capacity (Fig. 2a) that allow little scope for dispersal to generate synchrony because density differences are random, non-cyclic, and are continually created. Dispersal of only prey or predators produces results similar to those observed when both species disperse, emphasizing that the dispersal–predator interaction arises from phase-locking of predator–prey cycles, not from prey tracking predators that are themselves synchronized through predator dispersal (Supplementary Information).

In contrast to dispersal, the Moran effect does not generate phase-locking[2,4,26,27] and is slightly weakened by predator–prey cycles (Fig. 1a, b and Supplementary Information). The Moran effect operates by synchronously perturbing systems away from their intrinsically generated trajectories. In the presence of any source of spatially independent stochasticity, even perfectly synchronous environmental perturbations are unlikely to ensure that cycles emerge from the nadir in lock-step (Supplementary Information). Furthermore, the synchronizing effect of these perturbations does not increase with the density difference between the patches.

To test the model predictions we conducted a laboratory microcosm experiment. We grew the ciliate protist *Tetrahymena pyriformis* in paired batch cultures (habitat patches) using a system that closely matches model assumptions (Supplementary Information). We manipulated dispersal (no dispersal, or periodic exchanges of a small volume of culture medium between paired cultures), the Moran effect (spatially synchronous or spatially independent temperature fluctuations) and species interactions (absence or presence of the predatory ciliate *Euplotes patella*) in a full factorial design, as in the model simulations. This is the first factorial experiment manipulating all possible sources of spatial population synchrony.

The model proved capable of quantitatively predicting the experimental results (Fig. 1). As predicted, both dispersal and the Moran effect significantly increased the synchrony of *Tetrahymena* ($F_{1,36} = 6.5$, $P = 0.02$; $F_{1,36} = 20.3$, $P < 0.001$, respectively). Crucially, a significant interaction between dispersal and predators ($F_{1,36} = 4.8$, $P = 0.04$) indicated that dispersal was a much stronger synchronizing force in the presence of predators; dispersal had no significant effect on prey synchrony in the absence of predators ($F_{1,22} = 0.4$, $P = 0.5$). This interaction arose because predators decreased *Tetrahymena* synchrony in the absence of dispersal and increased it in the presence of dispersal, as predicted by the model. The remaining main effect and interaction terms were all nonsignificant, although we observed the predicted trend for predators to weaken the Moran effect (Fig. 1).

We found a strong qualitative match between the model predictions and experimentally observed treatment effect sizes across a large range of ecologically relevant parameter space (Fig. 3). Eighty-six per cent of Monte Carlo replicates predicted the three significant treatment effects found in our experiment (dispersal, the Moran effect, and the dispersal–predator interaction) to be the three strongest effects in the analysis of variance (ANOVA).

Our data rule out alternative mechanisms by which predators might enhance the synchronizing effect of dispersal. First, dispersal might synchronize predator dynamics, leading to increased prey synchrony because prey are impacted by fluctuations in predator density.
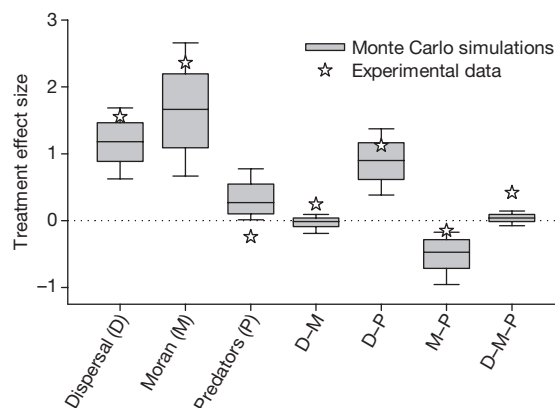
Figure 3 | The distribution of treatment effects generated by Monte Carlo simulation of 10,000 random parameterizations of the Rosenzweig–MacArthur model. Boxes represent the median, 25/75th, and 10/90th percentiles of the effect-size distribution and stars represent the experimental data. Treatment effects were calculated as the cumulative difference between, for example, dispersal against non-dispersal treatments (see Supplementary Information). Five model parameters were drawn from uniform intervals: attack rate (2.33, 5), dispersal rate (0.05, 0.25), process error standard deviation σ_ξ (0.05, 0.25), environment error standard deviation σ_ε (0.3, 0.9), and the correlation of the predator's and the prey's environment (−1, 1).

However, only the Moran effect significantly increased synchrony of the predator *Euplotes* (Fig. 4; $F_{1,16} = 13.6$, $P = 0.002$), suggesting that dispersal does not synchronize *Tetrahymena* indirectly by synchronizing the predator. Second, predators did not increase the strength of the Moran effect on prey synchrony, as would be expected if increased predator synchrony indirectly increased prey synchrony. Finally, lack of a significant correlation between prey and predator synchrony ($\rho = 0.05$, $t_{18} = 0.013$, $P = 0.99$) directly demonstrates that predator synchrony did not drive the synchrony of *Tetrahymena*.

In our experiment, *Euplotes* increased the temporal variability of *Tetrahymena* densities (coefficient of variation; $F_{1,84} = 49.6$, $P < 0.0001$). It is possible that increased temporal variability could increase spatial variability and thereby strengthen the synchronizing effect of dispersal, because the net prey dispersal rate is proportional to the difference in prey density between patches. However, our model demonstrates that simply increasing the variability of prey densities does not alter the impact of dispersal (Supplementary Information).

Instead, predators increased the synchronizing effect of dispersal by generating large-amplitude predator–prey cycles (Fig. 2d–f and Supplementary Fig. 7). The observed cycles were characterized by long periods with both predator and prey at low densities making them particularly susceptible to decoherence (Fig. 2e, f and Supplementary Information). However, in the presence of dispersal, cycles tended to emerge from the nadir nearly in lock-step, as predicted by the model (Fig. 2f). On average, paired prey populations emerged from the first cycle nadir (first non-zero prey sample density following first run of ≥2 consecutive samples with zero density) 1 day apart in the presence of dispersal and 2.5 days apart in the absence of dispersal, a significant difference (generalized linear model for Poisson-distributed data, $F_{1,18} = 5.68$, $P = 0.02$). In contrast, when predators were absent prey densities exhibited stochastic fluctuations around carrying capacity (Fig. 2d).

To test directly the ability of dispersal to synchronize emergence from the cycle nadir, we conducted a second, independent experiment (Supplementary Information). This experiment initiated replicate paired cultures slightly out of phase near the nadir of the predator–prey cycle, with half the replicates experiencing dispersal. As predicted by the model, in the absence of dispersal the difference in prey density between paired cultures increased as the 'leading' prey population emerged first from the cycle nadir and exhibited accelerating growth to high density. The prey density difference subsequently decreased as the leading prey population approached peak density and decelerated, allowing the 'trailing' prey population to (temporarily) catch up. Dispersal slowed prey growth in the leading patch and accelerated prey growth in the trailing patch, causing synchronous emergence from the cycle nadir and preventing the prey density difference from reaching large values. The results of this second experiment demonstrate that the model captures the detailed dynamical mechanisms linking dispersal to cycle synchrony in the first experiment.

Our results provide the first experimental demonstration of phase-locking of predator–prey cycles and extend the range of oscillatory dynamical systems in which phase-locking is known. Our results

support previous work suggesting that dispersal is a more effective synchronizing agent in the presence of cyclic dynamics[4,13,21,27].

Although the Moran effect significantly increased synchrony of both *Tetrahymena* and *Euplotes*, we found no evidence of environmental tracking in either species, or in our model simulations (Supplementary Fig. 8). Synchronous environmental forcing synchronizes population fluctuations by producing synchronous perturbations away from the intrinsically generated population trajectory. The Moran effect therefore can synchronize populations without population densities tracking environmental fluctuations[28]. Ciliates only noticeably track temperature fluctuations when maximum temperatures are sufficiently high to cause mass mortality[29].

In our experiment we found no significant effect of dispersal on *Euplotes* synchrony, in contrast to the model predictions (Fig. 4). This discrepancy may be due to lack of statistical power, as four replicates were lost as a result of predator extinction and *Euplotes* densities often were at or below the detection threshold for our sampling protocol.

Much previous work considers how coupling synchronizes oscillations[1]. Our work highlights the converse: oscillations enable coupling to produce synchrony. Our results provide a simple, general explanation for why many of the most dramatic examples of spatial synchrony in ecology comprise synchronized predator–prey oscillations. Like most examples of spatial synchrony, including Huygens' clocks, firing of the sinoatrial node in the mammalian heart, flashing of fireflies and light emission in lasers[1], it is the underlying oscillations that provide the scope for synchrony. Anthropogenic disturbances that alter or eliminate population cycles, such as vaccination against pathogens, climate change and species extirpations, may alter spatial synchrony and have unexpected consequences for species persistence.

METHODS SUMMARY

We conducted a fully crossed factorial experiment on the effect of dispersal, the Moran effect and predators on the synchrony of *T. pyriformis* dynamics in paired culture jars. Dispersal was accomplished by manual transfer of 10% of the experimental medium three times per week. We randomly varied the incubation temperature of culture jars on a daily basis between 20 and 30 °C according to prescribed schedules. Paired jars receiving the Moran effect experienced identical variability, whereas the remainder experienced independent variability. The predator treatment consisted of the addition of 15 individuals of *E. patella* to each jar in a pair on day 3 of the experiment. Each of the eight treatment combinations were replicated six times for a total of forty-eight experimental units. *Tetrahymena* and *Euplotes* densities were sampled each weekday using published methods[30], until day 64 of the experiment. We calculated the Pearson correlation of the first-order differences of *Tetrahymena* and *Euplotes* densities and conducted ANOVAs on the Fisher's z-transformed correlations.

We simulated a Rosenzweig–MacArthur predator–prey model in two patches and using the same eight treatment combinations as in the experiment. The experimental system matches key model assumptions (Supplementary Information). The equations were modified to include natural mortality in the prey, diffusive dispersal of individuals across patches, spatially correlated stochastic environmental fluctuations affecting background mortality rates, and process noise (a phenomenological description of other, spatially independent sources of random variation, including demographic stochasticity). We simulated the model for 2,048 time steps using a Runge–Kutta 4th/5th order algorithm, holding the stochastic parameters constant within each time step. We calculated synchrony of prey and predators in the manner described for the experiment, using only the latter 1,024 time-steps to ensure that initial transients had subsided. We conducted 10,000 Monte Carlo simulations in random multivariate parameter space to determine the robustness of our results (Supplementary Information).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.
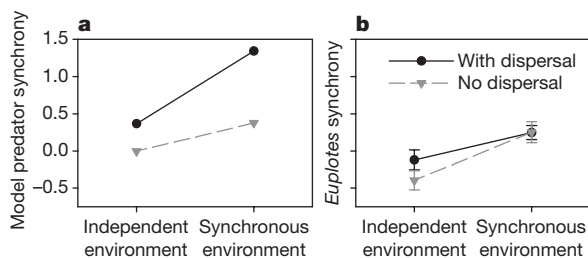
Figure 4 | Two-way interaction plot of the impact of dispersal and the Moran effect on predator synchrony. a, b, Results from our theoretical model (**a**) and our experimental microcosms (**b**) are shown. Points represent the mean z-transformed Pearson correlation (±1 s.e.m.). In the experiment, only the Moran effect significantly increases synchrony of the predator *Euplotes*.

1. Strogatz, S. H. *SYNC: How Order Emerges From Chaos in the Universe, Nature and Daily Life* (Hyperion, 2003).
2. Ranta, E., Kaitala, V., Lindström, J. & Linden, H. Synchrony in population dynamics. *Proc. R. Soc. Lond. B* **262**, 113–118 (1995).

1009

3.   Liebhold, A., Koenig, W. D. & Bjørnstad, O. N. Spatial synchrony in population dynamics. *Annu. Rev. Ecol. Evol. Syst* **35**, 467–490 (2004).

4.   Bjørnstad, O. N., Ims, R. A. & Lambin, X. Spatial population dynamics: analyzing patterns and processes of population synchrony. *Trends Ecol. Evol.* **14**, 427–432 (1999).

5.   Moran, P. A. P. The statistical analysis of the Canadian lynx cycle. *Aust. J. Zool.* **1**, 291–298 (1953).

6.   Grenfell, B. T. *et al.* Noise and determinism in synchronized sheep dynamics. *Nature* **394**, 674–677 (1998).

7.   Rohani, P., Earn, D. J. D. & Grenfell, B. T. Opposite patterns of synchrony in sympatric disease metapopulations. *Science* **286**, 968–971 (1999).

8.   Ims, R. A. & Andreassen, H. P. Spatial synchronization of vole population dynamics by predatory birds. *Nature* **408**, 194–196 (2000).

9.   Bjørnstad, O. N. Cycles and synchrony: two historical 'experiments' and one experience. *J. Anim. Ecol.* **69**, 869–873 (2000).

10. Peltonen, M., Liebhold, A. M., Bjørnstad, O. N. & Williams, D. W. Spatial synchrony in forest insect outbreaks: roles of regional stochasticity and dispersal. *Ecology* **83**, 3120–3129 (2002).

11. Heino, M., Kaitala, V., Ranta, E. & Lindström, J. Synchronous dynamics and rates of extinction in spatially structured populations. *Proc. R. Soc. Lond. B* **264**, 481–486 (1997).

12. Earn, D. J. D., Rohani, P. & Grenfell, B. T. Persistence, chaos and synchrony in ecology and epidemiology. *Proc. R. Soc. Lond. B* **265**, 7–10 (1998).

13. Blasius, B., Huppert, A. & Stone, L. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature* **399**, 354–359 (1999).

14. Rooney, N., McCann, K., Gellner, G. & Moore, J. C. Structural asymmetry and the stability of diverse food webs. *Nature* **442**, 265–269 (2006).

15. Vasseur, D. A. & Fox, J. W. Environmental fluctuations can stabilize food web dynamics by increasing synchrony. *Ecol. Lett.* **10**, 1066–1074 (2007).

16. Hudson, P. J. & Cattadori, I. M. The Moran effect: a cause of population synchrony. *Trends Ecol. Evol.* **14**, 1–2 (1999).

17. Ripa, J. Analysing the Moran effect and dispersal: their significance and interaction in synchronous population dynamics. *Oikos* **89**, 175–187 (2000).

18. Engen, S. & Sæther, B.-E. Generalizations of the Moran effect explaining spatial synchrony in population fluctuations. *Am. Nat.* **166**, 603–612 (2005).

19. Vasseur, D. A. Environmental colour intensifies the Moran effect when population dynamics are spatially heterogeneous. *Oikos* **116**, 1726–1736 (2007).

20. Ripa, J. & Ranta, E. Biological filtering of correlated environments: towards a generalised Moran theorem. *Oikos* **116**, 783–792 (2007).

21. Cazelles, B. & Boudjema, G. The Moran effect and phase synchronization in complex spatial community dynamics. *Am. Nat.* **157**, 670–676 (2001).

22. Cattadori, I. M., Haydon, D. T. & Hudson, P. J. Parasites and climate synchronize red grouse populations. *Nature* **433**, 737–741 (2005).

23. Abbott, K. C. Does the pattern of population synchrony through space reveal if the Moran effect is acting? *Oikos* **116**, 903–912 (2007).

24. Kendall, B. E. *et al.* Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. *Ecology* **80**, 1789–1805 (1999).

25. Hassell, M. P., Lawton, J. H. & Beddington, J. R. Components of arthropod predation. I. The prey death-rate. *J. Anim. Ecol.* **45**, 135–164 (1976).

26. Jansen, V. A. A. Phase locking: another cause of synchronicity in predator–prey systems. *Trends Ecol. Evol.* **14**, 278–279 (1999).

27. Jansen, V. A. A. The dynamics of two diffusively coupled predator–prey populations. *Theor. Popul. Biol.* **59**, 119–131 (2001).

28. Ranta, E., Lundberg, P., Kaitala, V. & Laakso, J. Visibility of the environmental noise modulating population dynamics. *Proc. R. Soc. Lond. B* **267**, 1851–1856 (2000).

29. Petchey, O. L. Environmental colour affects aspects of single-species population dynamics. *Proc. R. Soc. Lond. B* **267**, 747–754 (2000).

30. Fox, J. W. Testing a simple rule for dominance in resource competition. *Am. Nat.* **159**, 305–319 (2002).

Author Contributions Both authors conceived the experiment and analysed the results. D.A.V. conceived and analysed the model. Both authors wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.A.V. (david.vasseur@yale.edu).

METHODS

Experimental methods. Experimental units comprised pairs of 80 ml culture jars containing a solution of sterilized local spring water, $0.2\,\mathrm{g\,l^{-1}}$ of Protozoan Pellets (standardized pellets of crushed, dried plant matter produced by Carolina Biological Supply), and one wheat seed. Protozoan Pellets and wheat seeds provide energy and nutrients for bacteria, which are consumed by *Tetrahymena*. The eight treatment combinations were replicated six times, yielding $n = 48$ experimental units. Twenty-four hours before inoculation with *Tetrahymena*, three species of bacteria were added to each jar: *Bacillus subtilis*, *Enterobacter aerogenes* and *Bacillus cereus*; other unidentified bacteria were added with the protists. On day 0, *T. pyriformis* stock culture (1 ml) was added to each jar (stock culture density $5{,}858.7\,\mathrm{ml^{-1}}$). On day 3 of the experiment, 15 individuals per jar of *E. patella* were added to the 'predator' treatments by micropipette. Beginning on day 3, densities of *Tetrahymena* and *Euplotes* in small samples (~ 0.2 ml) were determined daily on weekdays using published methods[30]. The culture medium was refreshed weekly, beginning on day 4, by replacing 9 ml experimental medium with 10 ml fresh sterile medium; the difference accounted for loss due to sampling. The experiment continued until day 64, resulting in 44 observations per jar. Dispersal was accomplished by exchanging 8 ml medium between jars within a pair every Monday, Wednesday and Friday after sampling. This dispersal regime imposes density-independent (diffusive) dispersal characterized by equal per-capita dispersal rates for predators and prey.

Environmental fluctuations were accomplished daily by moving jars between 20 and 30 °C incubators according to prescribed schedules. Each schedule comprised 38 days at 20 °C and 26 days at 30 °C permuted by a random series possessing a reddened spectrum ($1/f^{0.5}$) (see ref. 19). This introduced a natural level of temporal autocorrelation into environmental fluctuations and ensured a consistent mean temperature for each jar (24.1 °C). We discarded and replaced two schedules prescribing incubation at 30 °C for more than 8 days in a 10 day period to better enable persistence of *Euplotes*. Jar pairs experiencing synchronized temperature fluctuations followed the same schedule; otherwise, jars followed independent schedules. Different replicate pairs within both synchronized and independent treatments experienced different time series of temperature fluctuations to avoid confounding environmental synchrony with the unique properties of any particular temperature time series.

We calculated the synchrony of *Tetrahymena* and *Euplotes* densities (x) within experimental units as the Pearson product-moment correlation (ρ) of the first-order differences, $r(t) = x(t+1) - x(t)$. We measured synchrony using cross-correlations rather than phase differences because our time series were too short to allow precise estimation of cycle phase. We measured synchrony in the same fashion in our theoretical model, so the experimental data directly test model predictions.

We conducted a three-factor ANOVA on the Fisher's z-transformed correlations, $z = 0.5 \ln[(1+\rho)/(1-\rho)]$, to test for differences in synchrony across treatments. Inspection of the $r(t)$ series revealed a highly synchronized initial (transient) decline in *Tetrahymena* across all jars. We removed the first 12 days of data from the analysis to ensure that this synchronous transient did not dominate our results. Removing more than 12 days of initial data did not qualitatively alter our results and produced only very minor quantitative changes. In four replicate pairs containing both *Tetrahymena* and *Euplotes*, densities dropped below our detection threshold in at least one jar and did not recover; these replicates were excluded from all analyses. Inspection of residuals for all statistical analyses indicated conformity with distributional assumptions.

Theoretical model. The modified Rosenzweig–MacArthur model of prey N and predators P in patch i is given by the non-dimensional model:

$$\frac{dN_i}{dt} = N_i(1 - N_i) - m_{N_i}N_i - \frac{aN_iP_i}{N_i + N_0} + d(N_j - N_i) + \varepsilon_{N_i}N_i$$

$$\frac{dP_i}{dt} = \frac{eaN_iP_i}{N_i + N_0} - m_{P_i}P_i + d(P_j - P_i) + \varepsilon_{P_i}P_i$$

where $i = 1, 2$ and $i \neq j$. Values for the attack rate $a = 3.0$, assimilation efficiency $e = 0.5$, and half-saturation density $N_0 = 0.3$ were chosen to generate predator–prey cycles that qualitatively mimic those seen in the experiment (Figs 1–3). Their exact values have little impact on the results, provided that they generate a limit cycle or damped oscillations in the absence of stochasticity (Supplementary Information).

Dispersal was controlled by the diffusive rate d and was equal to 0.15 (or 0) for both predators and prey. We imposed synchronized or independent fluctuations in the background per-capita mortality rates (m) according to $m_{N_i}(t) = 0.25 \exp[\xi_i(t)]$ and $m_{P_i}(t) = 0.5 \exp[\xi_i(t)]$ where ξ_i are discrete, normally distributed random variables with $\mu_\xi = 0$ and $\sigma_\xi = 0.6$ (see ref. 15 for rationale). In treatments without predators, P_i were initialized at zero, otherwise P_i and N_i were chosen randomly on the intervals (0.1, 0.2) and (0.5, 1.0), respectively.

Independent process error was imposed on each population by a discrete random normal variable ε with $\mu_\varepsilon = 0$ and $\sigma_\varepsilon = 0.15$. Our description of process error generates temporal variance in population size proportional to the square of mean population size, independent of spatially correlated environmental fluctuations and predator–prey cycles. This variance-mean scaling relationship is broadly consistent with theoretical arguments and empirical data[31].

31. Vasseur, D. A. & Gaedke, U. Spectral analysis unmasks synchronous and compensatory dynamics in phytoplankton communities. *Ecology* **88**, 2058–2071 (2007).

A highly annotated whole-genome sequence of a Korean individual

Jong-Il Kim[1,2,4,5]*, Young Seok Ju[1,2]*, Hansoo Park[1,5], Sheehyun Kim[4], Seonwook Lee[4], Jae-Hyuk Yi[1], Joann Mudge[6], Neil A. Miller[6], Dongwan Hong[1], Callum J. Bell[6], Hye-Sun Kim[4], In-Soon Chung[4], Woo-Chung Lee[4], Ji-Sun Lee[4], Seung-Hyun Seo[5], Ji-Young Yun[5], Hyun Nyun Woo[4], Heewook Lee[4], Dongwhan Suh[1,2,3], Seungbok Lee[1,2,3], Hyun-Jin Kim[1,3], Maryam Yavartanoo[1,2], Minhye Kwak[1,2], Ying Zheng[1,2], Mi Kyeong Lee[5], Hyunjun Park[1], Jeong Yeon Kim[1], Omer Gokcumen[7], Ryan E. Mills[7], Alexander Wait Zaranek[8], Joseph Thakuria[8], Xiaodi Wu[8], Ryan W. Kim[6], Jim J. Huntley[9], Shujun Luo[9], Gary P. Schroth[9], Thomas D. Wu[10], HyeRan Kim[4], Kap-Seok Yang[4], Woong-Yang Park[1,2,3], Hyungtae Kim[4], George M. Church[8], Charles Lee[7], Stephen F. Kingsmore[6] & Jeong-Sun Seo[1,2,3,4,5]

Recent advances in sequencing technologies have initiated an era of personal genome sequences. To date, human genome sequences have been reported for individuals with ancestry in three distinct geographical regions: a Yoruba African, two individuals of northwest European origin, and a person from China[1–4]. Here we provide a highly annotated, whole-genome sequence for a Korean individual, known as AK1. The genome of AK1 was determined by an exacting, combined approach that included whole-genome shotgun sequencing (27.8× coverage), targeted bacterial artificial chromosome sequencing, and high-resolution comparative genomic hybridization using custom microarrays featuring more than 24 million probes. Alignment to the NCBI reference, a composite of several ethnic clades[5,6], disclosed nearly 3.45 million single nucleotide polymorphisms (SNPs), including 10,162 non-synonymous SNPs, and 170,202 deletion or insertion polymorphisms (indels). SNP and indel densities were strongly correlated genome-wide. Applying very conservative criteria yielded highly reliable copy number variants for clinical considerations. Potential medical phenotypes were annotated for non-synonymous SNPs, coding domain indels, and structural variants. The integration of several human whole-genome sequences derived from several ethnic groups will assist in understanding genetic ancestry, migration patterns and population bottlenecks.

A bacterial artificial chromosome (BAC) library was constructed from AK1 genomic DNA. The genomic locations of about 100,000 AK1 BAC clones were ascertained by end-sequencing (Supplementary Table 1). Massively parallel DNA sequencing was performed using sequencing-by-synthesis with reversible-terminator chemistry on Illumina Genome Analyzers using two complementary strategies (Table 1, Supplementary Table 2 and Supplementary Fig. 1). First, selected genomic regions were sequenced at very high depth using overlapping BAC clones. Chromosome 20 was sequenced in this manner at 155× coverage, as were 390 other regions of the genome that are commonly affected by copy number variants (CNVs) (at an average

Table 1 | Overview of libraries and sequence data

Experiment type	Template DNA	Read length	Number of reads (M)	Total bases (Gb)	Number of aligned reads* (M)	Aligned bases* (Gb)	Percentage of genome covered	Coverage depth
Diploid								
SE short read	gDNA	1 × 36	519.5	18.7	393.3	14.2		
PE short read	gDNA	2 × 36	1,646.5	59.3	1,343.0	48.3	99.75	27.8×
PE long read	gDNA	2 × 88	123.3	10.9	95.4	8.4		
LIPE	gDNA	2 × 106	177.4	18.8	78.0	8.3		
Subtotal	gDNA		2,466.7	107.7	1,909.7	79.2		
Haploid								
BAC clones on Chr 20	742 BAC clones (targeted for Chr 20)	1 × 36, 2 × 36	299.8	10.8	187.4	6.7	96.3	155.3×
BAC clones on Chr 20	43 BAC clones (targeted for Chr 20)	2 × 36	96.7	3.5	77.1	2.8	94.6	531.7×
BAC clones on CNVRs	1,132 BAC clones (targeted for 390 CNVR)	2 × 36	234.1	8.4	130.6	4.7	95.1	47.8×
Subtotal	BAC clones		630.6	22.7	395.1	14.2		
Total	—	—	3,097.4	130.3	2,304.8	93.4	—	—

Chr, chromosome; CNVR, CNV region; gDNA, genomic DNA; LIPE, long insert-size paired end; PE, paired-end; SE, single-end.
* Sequences were aligned to the Human Genome Assembly build 36.3.

[1]Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul 110-799, Korea. [2]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, [3]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea. [4]Macrogen Inc., Seoul 153-023, Korea. [5]Psoma Therapeutics, Inc., Seoul 110-799, Korea. [6]National Center for Genome Resources, Santa Fe, New Mexico 87505, USA. [7]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [8]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [9]Illumina Inc., Hayward, California 94545, USA. [10]Department of Bioinformatics, Genentech Inc., South San Francisco, California 94080, USA.
*These authors contributed equally to this work.

of 151× coverage). Second, whole-genome sequencing was performed for the entire genome to an average depth of 27.8× using libraries of AK1 genomic DNA with different insert sizes to provide even coverage. Some sequences were generated using a reformulated cleavage reagent that removed thymine fluorophores more completely. This improved phasing and reduced background signals, error rates, and GC bias in longer reads (Supplementary Fig. 2), resulting in increased sequence yields and read lengths to 18 gigabases (Gb) per flow cell and 106 nucleotides, respectively. The average sequence quality was 24 (Q score[3]), and 74.4% of sequences aligned to the human genome reference (NCBI build 36.3) using the GSNAP alignment tool tolerating 5% mismatches[7–9]. A total of 99.8% of the reference genome was represented, and no coverage bias was appreciated apart from expected gaps at centromeres and other heterochromatic regions (Supplementary Discussion).

Bioinformatic filters were trained to detect and genotype SNPs in the aligned sequences. Filters ascertained by comparing SNP genotypes derived from sequencing with results from Illumina 370K genotyping array gave a positive predictive value and sensitivity of SNP detection of 99.9% and 95.0%, respectively (Supplementary Fig. 3 and Supplementary Table 3), and SNP genotype accuracy of 99.1% (Supplementary Fig. 4 and Supplementary Table 4). These filters detected 3,453,653 SNPs in the genome of AK1 (density of 1.21 per kilobase (kb)), of which 17.1% were new and 10,162 were non-synonymous (Supplementary Tables 5 and 6). These results were verified by hybridization of genomic DNA from AK1 to an Illumina 610K genotyping array, deep sequencing of chromosome 20 BAC clones, and Sanger resequencing of the AK1 genome (Supplementary Tables 7, 8 and Supplementary Discussion). The number of SNPs detected in the genome of AK1 was similar to that

of James Watson, higher than Craig Venter and the Chinese YH, and less than the Yoruba African, NA18507 (Fig. 1a, b and Supplementary Table 9), which may reflect differences in technical procedures or inter-individual variability[1–4]. Overlap among 9,527,824 SNPs detected in these five sequenced genomes indicated that 21% of AK1's SNPs were unique, and 8% were shared by all (Fig. 1b). A total of 2,110,403 AK1 SNPs were heterozygous, yielding a higher SNP diversity than in the Venter, Watson or YH genomes, but less than the Yoruba individual (heterozygous/homozygous SNP ratio of 1.57, and nucleotide diversity (π) of 7.40×10^{-4}; Supplementary Discussion). Sequencing of other genomes using uniform technical procedures is warranted to evaluate the proportion of genetic variance explained by differences within and between human populations.

We applied the same bioinformatic filters to the genome sequence of AK1 to detect indels (Supplementary Methods). The NCBI reference genome contained 7,910 exonic indel mismatches in comparison with the reference transcript (Supplementary Discussion). Excluding these, the SNP filters detected 170,202 indels (density of 0.060 per kb), of which 71,995 were homo- or hemizygous. Sixty-two per cent of indels were new and 55.9% were deletions (Supplementary Tables 10 and 11). The size range detected was −29 to +5 nucleotides, with approximately normal frequency distribution (Supplementary Fig. 5). Two-hundred-and-twelve AK1 indels mapped to coding domains, which was three times greater than that reported for the YH genome and one-fourth of that reported for the Venter genome[1,4] (Supplementary Fig. 6 and Supplementary Table 12). These marked differences reflect substantial differences either between individuals or between technical procedures, highlighting the need for definition of foundational data standards. Indel results
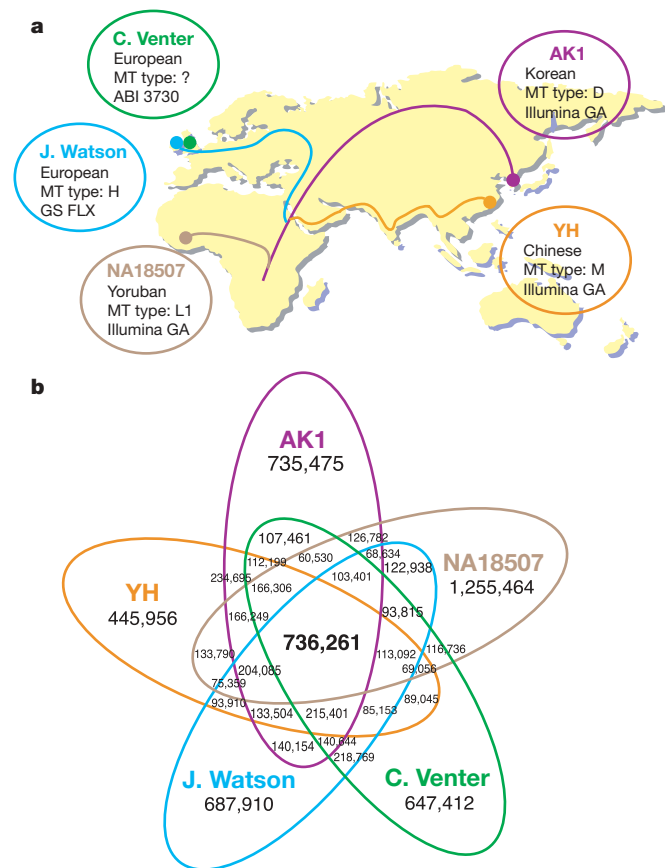
Figure 1 | Geographic map and Venn diagram of five sequenced genomes, indel distribution, and SNP-indel densities correlation. a, Geographic map showing the regions of ancestry of five sequenced genomes. MT type, mitochondrial haplogroup. **b,** The number of SNPs overlapping between five genomes. **c,** Correlation between SNP–indel densities on chromosome 6 (per 10-kb window). From top: SNP density, indel density, SNP–indel density (moving average of ten 10-kb windows), SNP density in a portion of chromosome 6, and indel density along the same portion of chromosome 6. The *x* axis represents the nucleotide position in Mb.
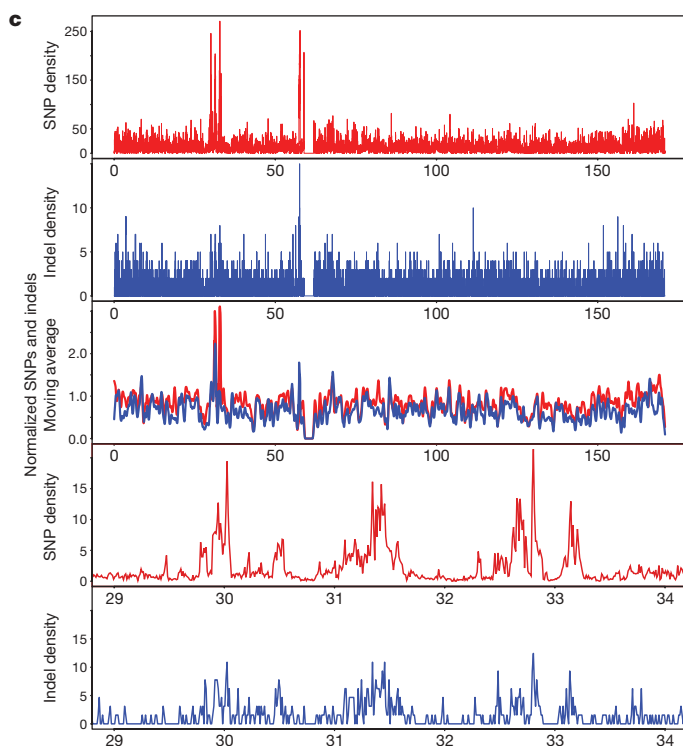
1012

were confirmed by Sanger resequencing of AK1 genomic DNA and deep sequencing of chromosome 20 BAC clones, showing that the sensitivity of indel detection was less than 80%, whereas the positive predictive value was 100% (Supplementary Discussion). Indel underestimation was unavoidable in local repetitive or homopolymeric sequences containing indels at, or near, the ends of reads (Supplementary Discussion). Seventy coding-domain indels were homozygous, of which 26 were in genes with Online Mendelian

Inheritance in Man (OMIM) entries, 13 of which had medical phenotypes (Supplementary Table 13).

Highly significant pairwise correlations of SNP and indel densities were observed throughout the genome (Pearson's correlation[10] was 0.40 genome-wide, $P < 10^{-300}$; Fig. 1c, Supplementary Fig. 7 and Supplementary Table 14). This SNP–indel correlation seems to be a general phenomenon in individual human genomes, rather than a technical artefact, because it was also detected in the YH genome
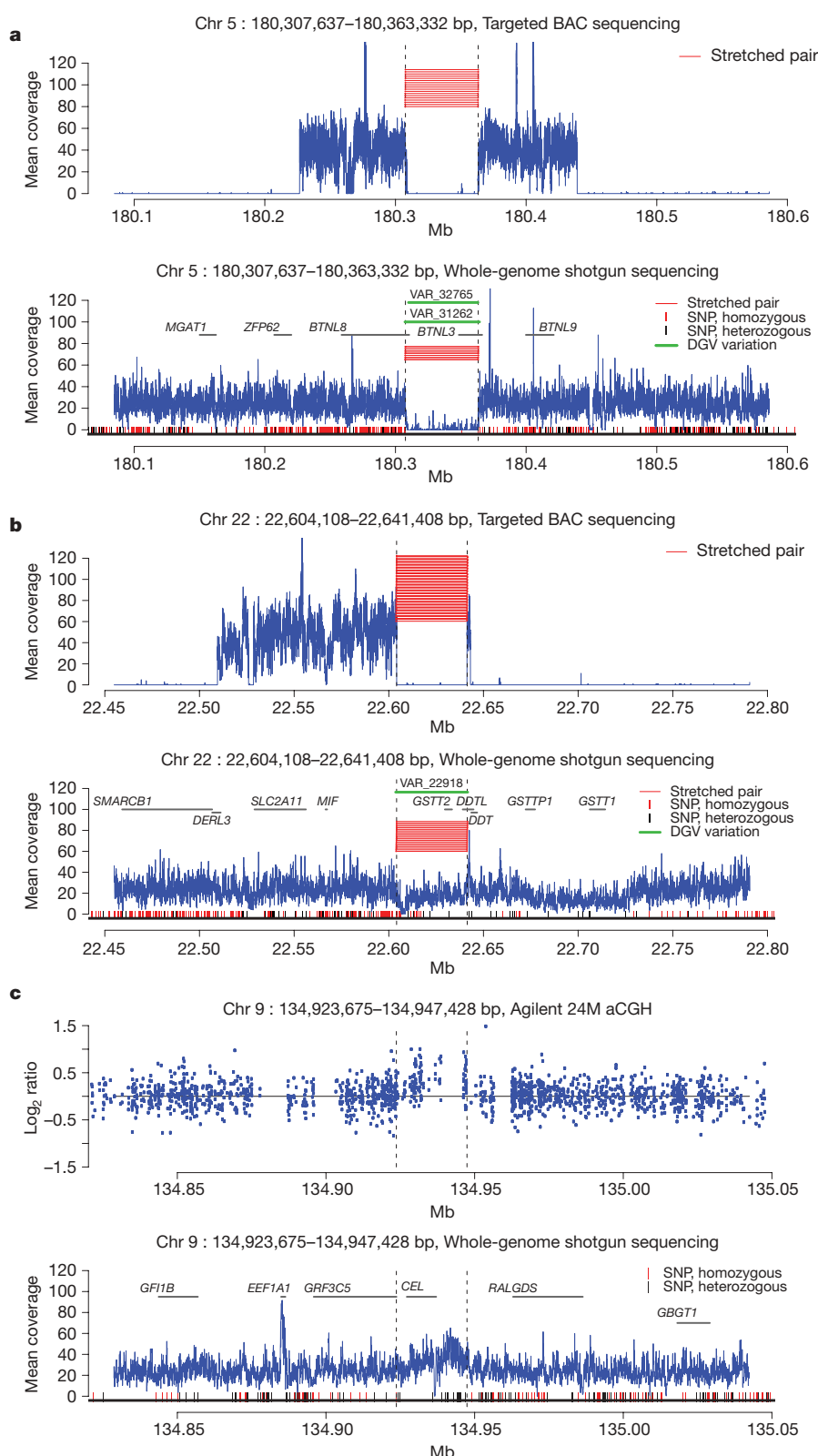
Figure 2 | Representative examples of genomic variations in AK1. a, Homozygous deletion identified by targeted haploid sequencing (top) and diploid sequencing (bottom). Stretched sequencing pairs and a drop in sequencing coverage define the deletion in both panels. Chr, chromosome. **b,** Heterozygous deletion identified by targeted haploid sequencing (top) and confirmed by diploid sequencing (bottom). Stretched pairs confirm the deletion in the diploid sequence but complete coverage drop is not detected. **c,** Copy number gain is identified by CGH microarray (top) and confirmed by increased coverage for the corresponding genomic region by diploid sequencing (bottom). For all panels: blue, fold coverage; horizontal red lines, stretched sequence pairs; green, CNV region in the DGV; grey, gene; vertical red bars, homozygous SNPs; vertical black bars, heterozygous SNPs; and broken vertical grey lines define the boundaries of the structural variants.

(Supplementary Table 14) and has been reported for other eukaryotes, including primates[11–16]. SNP–indel density covariation was not a function of coverage depth or gene density (Supplementary Table 14). Genome-wide correlation of SNP and indel densities in individual human genomes is a new finding, and suggests that unifying molecular or temporal considerations underpin the generation and/or removal of both types of variants[11–16].

Several, complementary approaches were used for CNV detection in AK1 (Supplementary Fig. 8). For large deletions of the AK1 genome, we used events identified in deep sequencing of 1,132 BAC clones as a training set (Supplementary Fig. 9). Most showed reduced coverage, predominance of homozygous SNPs (for heterozygous deletions), and alignment of 'stretched' paired-end reads in whole-genome sequences (Supplementary Fig. 10). We used these criteria to filter candidate deletions detected by a new, custom-designed 24 million probe set array comparative genomic hybridization (CGH), which found 1,237 CNV regions in total (Supplementary Fig. 11 and Supplementary Table 15), as well as genotyping microarrays (Supplementary Table 16). Figure 2a shows an example of a deletion in the genome of AK1 detected both by BAC sequencing and by CGH with the characteristics mentioned above. Figure 2b shows another deletion where the coverage drop in whole-genome sequencing was not as prominent as in BAC sequencing. However, the latter region contains all other features of deletions, illustrating the usefulness of BAC sequencing for CNV detection. The 238 regions that met these conservative criteria represent the most reliable list of true deletions yet identified in an individual genome sequence (Supplementary Fig. 10 and Supplementary Table 17). Deletions in AK1 ranged from 277 to 196,900 bases in length, and totalled 2.4 Mb. One-hundred-and-forty-eight of these had not previously been described in the Database of Genomic Variants as of 10 November 2008 (DGV; http://projects.tcag.ca/variation/).

Copy number gains in AK1 were selected conservatively with three approaches for different sizes of insertions: (1) array CGH described above yielded insertions ranging from 2.15 kb to 1.06 Mb, (2) aligned end-sequences of BACs yielded insertions of 16.8 to 357.1 kb, and (3) aligned long-insert paired-end reads yielded insertions of 0.9 to 2.2 kb (Supplementary Tables 18–20). These regions were confirmed by increased sequencing coverage of these genomic regions (Supplementary Figs 12–14). An example is shown in Fig. 2c, in which an increased signal on a microarray coincided with a corresponding significant gain of sequencing coverage. The AK1 genome contained 77 copy number gains, totalling 7.0 Mb. Thirty-three (42.8%) of these were absent from the DGV and were therefore considered to be new.

Non-synonymous SNPs detected in AK1 were compared with those identified in the YH and the Yoruban genomes[3,4] (which were ascertained using technical approaches similar to those used here) (Fig. 3a). Although only 37% of AK1 SNPs were shared among these three genomes, 57% of genes that contained non-synonymous SNPs in AK1 were common to all three (Fig. 3a and Supplementary Discussion). These data indicate that a subset of genes is enriched for non-synonymous SNPs in these individuals. Ontology analysis of this gene subset showed enrichment for functions associated with environmental adaptation, such as sensory function, immunological function, and signal transduction (Supplementary Table 21). Possibly, these genes have heightened diversity and/or many pseudogenes.

Using Trait-o-matic—an algorithm for high-throughput variant annotation—773 SNPs that were potentially associated with clinical phenotypes were identified (J. V. Thakuria and G. M. Church, manuscript in preparation; Supplementary Table 22). Of these, 269 were relatively common SNPs previously associated with risk of complex disorders or traits. For example, the genome of AK1 contained 90 SNPs that have shown associations with susceptibility to various cancers, 34 SNPs with type II diabetes mellitus, 13 with Alzheimer's
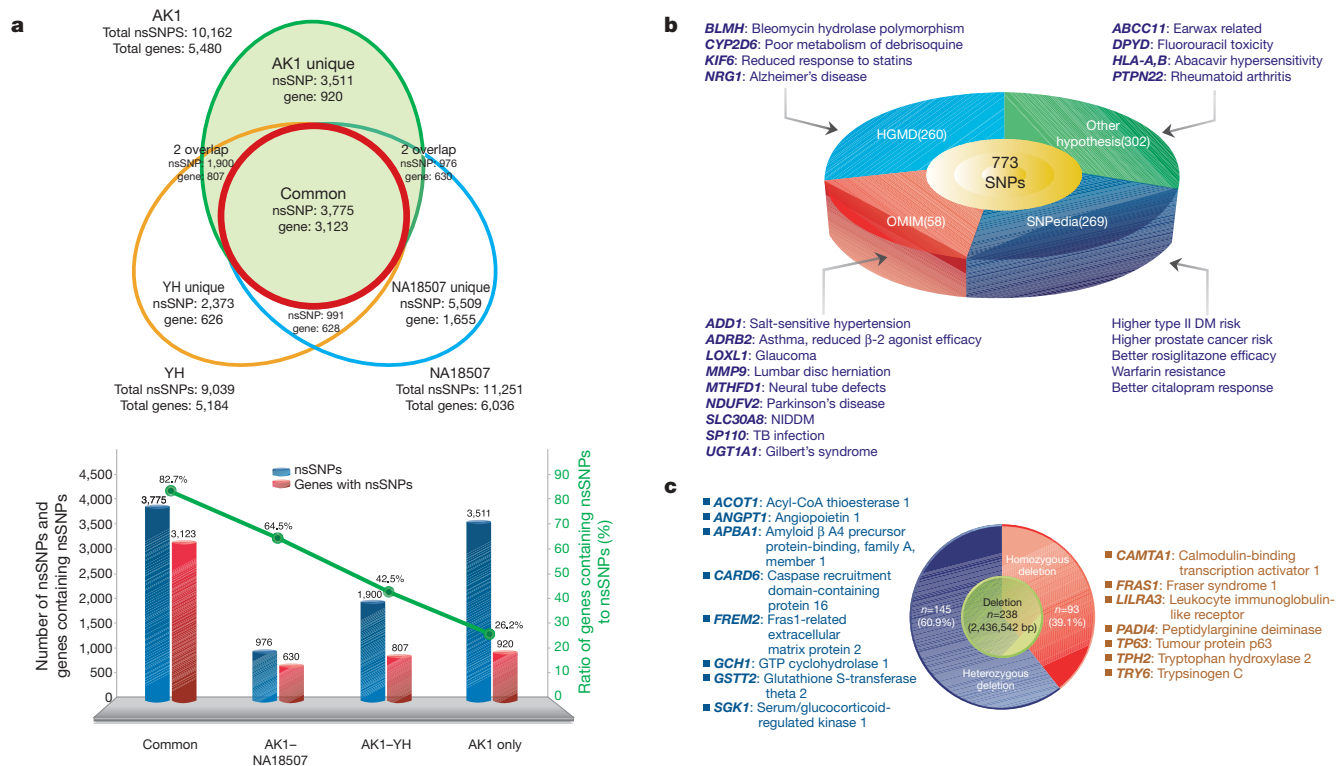
Figure 3 | Potential implications of AK1 variants and comparisons of non-synonymous SNPs among three sequenced genomes. a, Top, the numbers of non-synonymous SNPs (nsSNPs) and genes containing non-synonymous SNPs are compared between the Korean (AK1), Han Chinese (YH) and Yoruban (NA18507) genomes. Bottom, comparison of non-synonymous SNPs and genes containing non-synonymous SNPs in AK1 with those in the YH and Yoruban genomes. Common denotes shared by three genomes. Left axis: number of nsSNPs (blue) or genes containing nsSNPs (red); right axis: ratio (%) of the number of nsSNP genes to the number of nsSNPs (green). b, Seven-hundred-and-seventy-three SNPs potentially associated with clinical phenotypes derived from the database of human gene mutation data (HGMD), OMIM, SNPedia and other hypotheses. DM, diabetes mellitus; NIDDM, non-insulin-dependent diabetes mellitus; TB, tuberculosis. c, Genes affected by large homozygous and heterozygous chromosomal deletions.

disease, and seven with rheumatoid arthritis. These data should be interpreted cautiously, however, because risk factors for complex diseases, for example, rheumatoid arthritis, differ in northwest European and Korean populations[17], and because the translation of genetic burden into risk assessment for polygenic traits is rudimentary. The genome of AK1 also contained 504 non-synonymous SNPs in genes associated with complex or Mendelian disorders or traits. Of these, 22 were stop codons and five were homozygous. Among Mendelian traits, AK1 was homozygous for a variant conferring dry earwax[18] that has a high allele frequency in Koreans. Eighteen variants of pharmacogenetic relevance were identified, potentially affecting dosing, efficacy and/or toxicity of β-2-adrenoceptor agonists, statins, rosiglitazone, warfarin, citalopram, abacavir, debrisoquine, bleomycin, fluorouracil and aramycin-C (Fig. 3b).

CNVs have shown associations with common, complex disorders in humans. One-hundred-and-six genes were affected by CNV losses in AK1 (Fig. 3c). One gene deleted in the genome of AK1 was leukocyte immunoglobulin-like receptor (*LILRA3*). Most northeast Asians have functional loss or deletion of this locus, which has been suggested to be under positive or balancing selection[19].

We have obtained the genome sequence of a Korean individual by a unique combination of whole-genome shotgun sequencing, targeted BAC sequencing, and custom-designed high-resolution array CGH. This combination of approaches improved the accuracy of SNP, indel and CNV detection, and will assist in the assembly of contiguous sequences. Agreement on technical standards for individual genome sequences will aid in comparisons between genomes and, ultimately, to associations with phenotypic differences.

METHODS SUMMARY

Genomic DNA samples were obtained from an anonymous healthy Korean adult male (AK1) with normal karyotype (Supplementary Fig. 15), using guidelines approved by the Institutional Review Board of Seoul National University (approval C-0806-023-246). AK1 provided written consent for public release of genomic data. A BAC clone library was prepared using standard methods[20]. End sequences of 96,768 BAC clones were generated with an ABI 3730xl DNA analyser. A minimally overlapping BAC tiling path of chromosome 20, and 1,132 BAC clones from 390 common CNV regions were selected for haploid targeted sequencing. Genome-wide genotyping and CNV detection were performed using Agilent custom 24 million feature CGH array set, as well as Illumina Human cnv370- and 610-quad Beadchip. Short and long insert paired-end read libraries were generated from pooled BAC clone DNA or genomic DNA as described[3]. Paired-end and singleton, 36–106 nucleotide reads were generated using Illumina Genome Analyser (GA) and GAII instruments as described[3]. Long reads were obtained by use of multiple 36 nucleotide sequencing kits with reformulated cleavage reagent provided by Illumina for evaluation. This reformulated reagent has now been made available in all new Illumina sequencing-by-synthesis (SBS) reagent kits. The total sequencing cost took less than 200,000 US dollars with total run time of 6 weeks using three GA instruments. Sequences and runs were used in analyses if the average Q scores were ≥20. Sequences were aligned to NCBI build 36.3 using GSNAP[21]. SNPs and indels were identified using optimized filters through Alpheus[8]. Custom scripts were developed to identify CNVs in short, paired-end reads by the occurrence of clusters of reads with aligned insert sizes deviating from the mean by >2 standard deviations, and based on contiguous regions with significantly increased or depressed coverage. Putative SNPs, indels and deletions were validated by targeted Sanger sequencing (Supplementary Fig. 16 and Supplementary Table 23). Variants that were previously associated with a clinical phenotype or risk in other studies were identified with Trait-o-matic. Statistical analysis was performed using JMP-Genomics (SAS Institute) or R (http://www.R-project.org).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 March; accepted 18 June 2009.
Published online 8 July 2009.

1. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* 5, e254 (2007).
2. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876 (2008).
3. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59 (2008).
4. Wang, J. *et al.* The diploid genome sequence of an Asian individual. *Nature* 456, 60–65 (2008).
5. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* 409, 860–921 (2001).
6. Venter, J. C. *et al.* The sequence of the human genome. *Science* 291, 1304–1351 (2001).
7. Sugarbaker, D. J. *et al.* Transcriptome sequencing of malignant pleural mesothelioma tumors. *Proc. Natl Acad. Sci. USA* 105, 3521–3526 (2008).
8. Miller, N. A. *et al.* Management of high-throughput DNA sequencing projects: alpheus. *J. Comput. Sci. Syst. Biol.* 1, 132–148 (2008).
9. Mudge, J. *et al.* Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PLoS One* 3, e3625 (2008).
10. Sokal, R. R. & Rohlf, F. J. *Biometry: the Principles and Practice of Statistics in Biological Research* 3rd edn (W.H. Freeman, 1995).
11. Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* 21, 12–27 (2003).
12. Hardison, R. C. *et al.* Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res.* 13, 13–26 (2003).
13. Petrov, D. A. DNA loss and evolution of genome size in *Drosophila. Genetica* 115, 81–91 (2002).
14. Longman-Jacobsen, N., Williamson, J. F., Dawkins, R. L. & Gaudieri, S. In polymorphic genomic regions indels cluster with nucleotide polymorphism: Quantum Genomics. *Gene* 312, 257–261 (2003).
15. Silva, J. C. & Kondrashov, A. S. Patterns in spontaneous mutation revealed by human-baboon sequence comparison. *Trends Genet.* 18, 544–547 (2002).
16. Tian, D. *et al.* Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature* 455, 105–108 (2008).
17. Lee, H. S. *et al.* Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations. *Arthritis Rheum.* 60, 364–371 (2009).
18. Yoshiura, K. *et al.* A SNP in the *ABCC11* gene is the determinant of human earwax type. *Nature Genet.* 38, 324–330 (2006).
19. Hirayasu, K. *et al.* Evidence for natural selection on leukocyte immunoglobulin-like receptors for HLA class I in Northeast Asians. *Am. J. Hum. Genet.* 82, 1075–1083 (2008).
20. Frijters, A. C. J. *et al.* Construction of a bacterial artificial chromosome library containing large EcoRI and HindIII genomic fragments of lettuce. *Theor. Appl. Genet.* 94, 390–399 (1997).
21. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875 (2005).
22. Chen, R., Sodergren, E., Weinstock, G. M. & Gibbs, R. A. Dynamic building of a BAC clone tiling path for the Rat Genome Sequencing Project. *Genome Res.* 14, 679–684 (2004).

Author Contributions J.-S.S. managed the project. J.-S.S., W.-Y.P., H.K., H.K., H.-S.K., I.-S.C., W.-C.L. and J.-S.L. prepared the BAC library. Seon.L., S.-H.S., J.-Y.Y., H.N.W., R.W.K., J.J.H., Sh.L., G.P.S., H.K., K.-S.Y. and H.K. executed sequencing of the genome. J.-I.K., Y.S.J., S.K., J.M., N.A.M., D.H., C.J.B., H.L., H.-J.K., M.K.L. and T.D.W. analysed sequence data. Y.S.J., Ha.P., D.S., Seung.L., M.Y., M.K., Y.Z., Hy.P. and J.Y.K. validated DNA sequence variations. Ha.P., O.G., R.E.M. and C.L. were involved with microarray experiments. A.W.Z., J.T., X.W. and G.M.C. performed Trait-o-matic analysis. J.-S.S., J.-I.K., Y.S.J., J.-H.Y. and S.F.K. wrote the manuscript.

Author Information Data have been deposited in the NCBI short read archive under accession number SRA008370. These data are also available freely from http://gmi.ac.kr. SNPs and indels are deposited in the dbSNP database under handle GMI. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.-S.S. (jeongsun@snu.ac.kr).

METHODS

Sample preparation. Sperm samples and peripheral blood were obtained from an anonymous healthy Korean adult male in his thirties, following the guidelines by the Institutional Review Board of Seoul National University (approval number C-0806-023-246). High molecular mass genomic DNA was extracted from sperm and peripheral blood using standard protocols.

BAC clone library preparation and end sequencing. The BAC library was prepared from a partial HindIII digest of high molecular weight genomic DNA from sperm in the vector pECBAC1 using standard methods[20]. BAC clone DNA was extracted from a single colony by a standard alkaline lysis extraction method. Inserts of BAC clones were sequenced from both ends using universal primers, an ABI 3730xl DNA analyser and an ABI BigDye Terminator cycle sequencing (Applied Biosystems).

High-depth coverage data of single chromosome and frequent CNV region (CNVR) was obtained using targeted BAC GA sequencing. Minimally overlapping BAC clone tiling path of chromosome 20 was constructed using BAC end sequences[22]. Also, we selected 390 CNV regions from DGV at which the frequency of variants among the population was more than 20% (including genes). From the CNVRs, 1,132 BACs were selected. BAC clones were amplified and DNA was extracted for GA library preparation.

Genome Analyser library preparation. We constructed libraries for massive parallel sequencing under the manufacturer's standard protocol (Illumina, Inc.)[3].

Genome sequencing. Paired-end and singleton 36 nucleotide-long reads were generated using Illumina Genome Analyser (GA) and GAII instruments using standard protocols[3].

Long read (2 × 88 bp, 2 × 106 bp) GA sequencing. Longer reads were enabled by an improved cleavage mix increasing the efficiency of thiamine fluorophore cleavage, thereby decreasing background noise. Paired 88-bp reads and 106-bp reads were generated with improved cleavage mix on Illumina GA II instruments using standard protocols[3]. Sequences passing standard Illumina GA pipeline filters (v1.1) were retained. Long insert paired reads were trimmed to 50 nucleotides for insertion detection.

Sequence alignments. Short-reads were aligned using the GSNAP[21] alignment tool (Supplementary Methods).

Validation studies. Sixty-five SNPs, 60 indels and ten large deletions were selected for Sanger validation (Supplementary Discussion). PCR was performed in 50 μl containing 50 ng genomic DNA, 10 pmol of forward and reverse primer each, using Ex Taq (Takara), at 95 °C for 10 min, 35 cycles of 95 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s, and finally 72 °C for 10 min, using genomic DNA (Supplementary Table 24). PCR products were subsequently purified with AccuPrep PCR purification kit (BioNeer). Sanger sequencing was performed using an ABI 3730xl DNA analyser and ABI BigDye Terminator cycle sequencing.

SNP and indel sensitivity. To estimate the sensitivity of our SNP and indel detection, we used high-depth GA data from 742 BAC clones targeting chromosome 20. We selected 22.255 Mb of chromosome 20 with single BAC coverage for the identification of highly reliable SNPs and indels. Filter conditions were identical as homozygous SNP and indel detection used in autosomes. We estimated the sensitivity by comparing them with SNPs and indels identified by whole-genome shotgun data (Supplementary Discussion).

SNP and indel correlation. Binned counts observed for both SNPs and indel start positions (bin size = 10 kb) were computed, and pairwise correlations were calculated. Bins with no coverage were removed from the analysis. To investigate the potential confounding effect of coverage and gene density on the sample SNP–indel correlation, we computed the partial correlation[10] between SNP and indel counts, correcting for the effects of coverage variation and/or gene density (bin size = 10 kb). All pairwise correlations and partial correlations were computed using pcor and pcor.test from the ggm library in R. Moving average plots were generated in R using the lowess (locally weighted robust scatterplot smoothing) smooting function. Each 10-kb window was smoothed using the 100 surrounding windows. Lowess calculations were done with three iterations and without any linear interpolation of adjacent points.

Detection of true positive deletions with paired end GA data from 1,132 BACs. A sequence read pair with an insert size of more than 500 bp was defined as a stretched pair. Two criteria were used to finalize deletion list: the presence of 15 or more stretched pairs with the reads of either side clustering within 500-bp region, and the coverage of the candidate region between the clusters of stretched-pair reads less than half of those in the flanking regions of both sides. See Supplementary Methods for more details.

Detection of structural variations using microarrays. Putative structural variations were detected with microarrays (Illumina BeadChip 370K, 610K, and Agilent 24M aCGH; Supplementary Methods). For Illumina BeadChips, normalized bead intensity data and genotype calls were obtained with Illumina BeadStudio 3.1 software. Results from Agilent 24M aCGH were analysed on Nexus software (BioDiscovery Inc.). Each aberration call was manually checked to confirm the accuracy of the calls.

Detection of amplification with BAC end sequencing. BACs less than 40 kb were considered as significantly short. The co-localization of two or more short BACs was considered as the candidate region of amplification.

Detection of amplification with LIPE. Compressed pair was defined when it had a shorter than 2,000 bp insert size. The criteria used to finalize amplification list was the occurrence of ten or more de-redundant compressed pairs with the reads of either side clustering within 4 kb region. See Supplementary Methods for more details.

Confirmation of structural variants on diploid GA sequencing. The putative structural variants were confirmed using diploid GA. To confirm reliable large deletions, three criteria had to be met: (1) relative coverage drops compared to those of the flanking regions; (2) the existence of stretched paired-end sequence; and (3) the loss of heterozygous SNPs under the regions. Relative increase of sequencing coverage was used as a confirmation tool for copy number gains.

High-throughput annotation; Trait-o-matic. Trait-o-matic system was used for annotating 3.45 million SNPs of AK1 (Supplementary Methods). A working reference is available at: http://wiki.github.com/xwu/trait-o-matic/download-installation.

LETTERS

Neurotransmission selectively regulates synapse formation in parallel circuits *in vivo*

Daniel Kerschensteiner[1,2]†, Josh L. Morgan[1,2], Edward D. Parker[3], Renate M. Lewis[1] & Rachel O. L. Wong[1,2]

Activity is thought to guide the patterning of synaptic connections in the developing nervous system. Specifically, differences in the activity of converging inputs are thought to cause the elimination of synapses from less active inputs and increase connectivity with more active inputs[1,2]. Here we present findings that challenge the generality of this notion and offer a new view of the role of activity in synapse development. To imbalance neurotransmission from different sets of inputs *in vivo*, we generated transgenic mice in which ON but not OFF types of bipolar cells in the retina express tetanus toxin (TeNT). During development, retinal ganglion cells (RGCs) select between ON and OFF bipolar cell inputs (ON or OFF RGCs) or establish a similar number of synapses with both on separate dendritic arborizations (ON-OFF RGCs). In TeNT retinas, ON RGCs correctly selected the silenced ON bipolar cell inputs over the transmitting OFF bipolar cells, but were connected with them through fewer synapses at maturity. Time-lapse imaging revealed that this was caused by a reduced rate of synapse formation rather than an increase in synapse elimination. Similarly, TeNT-expressing ON bipolar cell axons generated fewer presynaptic active zones. The remaining active zones often recruited multiple, instead of single, synaptic ribbons. ON-OFF RGCs in TeNT mice maintained convergence of ON and OFF bipolar cells inputs and had fewer synapses on their ON arbor without changes to OFF arbor synapses. Our results reveal an unexpected and remarkably selective role for activity in circuit development *in vivo*, regulating synapse formation but not elimination, affecting synapse number but not dendritic or axonal patterning, and mediating independently the refinement of connections from parallel (ON and OFF) processing streams even where they converge onto the same postsynaptic cell.

ON bipolar cells, which depolarize in response to light, receive and invert photoreceptor signals through metabotropic glutamate receptors (mGluR6) on their dendrites. OFF bipolar cells, on the other hand, express ionotropic glutamate receptors and hyperpolarize upon illumination[3,4]. Thus, the parallel processing of light increment (ON) and decrement (OFF) signals that is central to visual system function is initiated at the first synapse in the retina. To create an imbalance of glutamate release from ON and OFF bipolar cells in the inner plexiform layer where their axons relay signals to RGCs, we generated transgenic mice in which a promoter fragment of mGluR6 (ref. 5) drove expression of the light chain of TeNT and yellow fluorescent protein (*mGluR6-YFP/ TeNT*, Fig. 1a). TeNT, a bacterial protease, cleaves vesicle-associated membrane protein 2 (VAMP2) and inhibits vesicle fusion[6]. In *mGluR6-YFP/TeNT* mice, transgene expression was limited to ON bipolar cells which ramify their axons in the inner half of the inner plexiform layer (Fig. 1b), and was present throughout the period of bipolar cell–RGC synaptogenesis (Supplementary Fig. 1). We tested the proteolytic activity of TeNT using an antibody against VAMP2

(ref. 7). Double-labelling for the vesicular glutamate transporter 1 (VGluT1), which is present in all bipolar cell axon terminals, verified that VAMP2 was selectively depleted from ON bipolar cells (Fig. 1c–h). To examine glutamate release, we recorded spontaneous excitatory postsynaptic currents (sEPSCs) from large monostratified RGCs in retinal flat mount preparations (Fig. 1i). ON RGCs from *mGluR6-YFP/ TeNT* mice showed very few and lower amplitude sEPSCs compared to wild type, whereas OFF RGC sEPSCs were normal in frequency and amplitude (Fig. 1i–k). Focal application of the glutamate receptor agonist kainate (100 μM) near the soma of ON RGCs elicited excitatory currents of similar amplitude and without failure in wild-type and *mGluR6-YFP/TeNT* mice (Supplementary Fig. 2), showing that the reduction in spontaneous events probably reflects a blockade of presynaptic transmitter release.

To examine ON and OFF responses, which are driven by glutamate release from bipolar cells, we used multi-electrode arrays to record from ensembles of RGCs in wild-type and *mGluR6-YFP/TeNT* mice during visual stimulation. Receptive fields were characterized by the spike-triggered average (STA) stimulus during white noise presentation[8]. In wild-type retinas, all RGCs were light responsive, and ON and OFF RGCs displayed similar receptive field size and sensitivity (Fig. 1l–o). By contrast, in *mGluR6-YFP/TeNT* mice OFF RGCs showed normal receptive field properties, but ON responses were greatly reduced or absent (Fig. 1l–o). Taken together, these results suggest that in *mGluR6-YFP/TeNT* mice both spontaneous and evoked glutamate release were diminished from ON but not OFF bipolar cells.

Early in development, monostratified RGCs elaborate dendrites throughout the depth of the inner plexiform layer[9,10] and appear to be contacted by ON and OFF bipolar cells[11]. With maturation, dendritic branches are pruned and a single narrowly stratified arborization connected to either ON or OFF bipolar cells is maintained[12]. To visualize dendrites and excitatory synapses, we biolistically labelled single RGCs with tdTomato (tandem dimer Tomato) and PSD95– CFP (postsynaptic density protein 95–cyan fluorescent protein)[13]. Figure 2 shows that, despite the selective reduction of input activity from ON bipolar cells in *mGluR6-YFP/TeNT* mice, ON RGC dendrites stratified normally and their lateral branching patterns were indistinguishable from wild-type littermates (Fig. 2a, b, e and f and Supplementary Fig. 3). Excitatory postsynaptic sites marked by PSD95– CFP were localized precisely to appositions on ON RGC dendrites with TeNT-expressing ON bipolar cells (Fig. 2c, Supplementary Fig. 4 and Supplementary Movie 1), but were reduced in density by ~50% compared to wild type at postnatal day 21 (Fig. 2d, g and h). The remaining synapses showed similar centroperipheral distributions across dendritic fields as in wild type (Fig. 2d, i and j). These results indicate that input selection and dendritic stratification of postsynaptic RGCs in mice occur independently of bipolar cell activity

[1]Department of Anatomy and Neurobiology, Washington University School of Medicine, St Louis, Missouri 63110, USA. [2]Department of Biological Structure, [3]Department of Ophthalmology, University of Washington, Seattle, Washington 98195, USA.
†Present Address: Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, St Louis, Missouri 63110, USA.
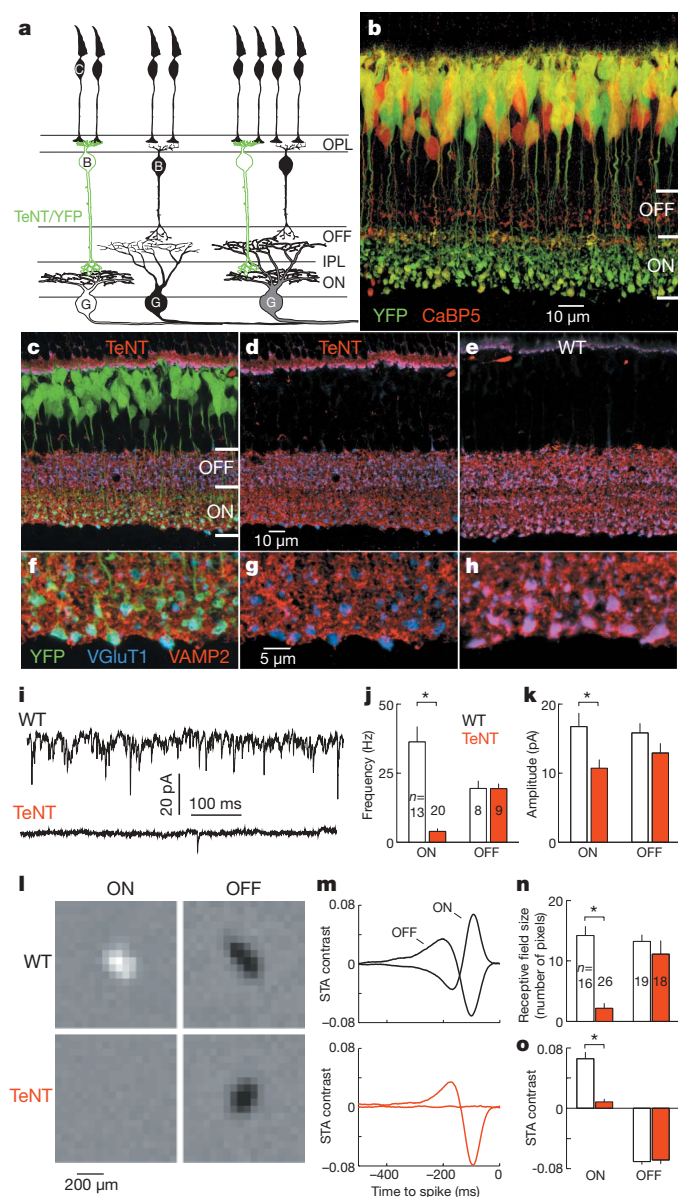
Figure 1 | Selective blockade of glutamate release from retinal ON bipolar cells. a, ON and OFF pathways. B, bipolar cell; C, cone; G, ganglion cell; IPL, inner plexiform layer; OPL, outer plexiform layer. ON (white), OFF (black) and ON-OFF responsiveness (grey) indicated by fill colour. **b,** YFP expression in *mGluR6-YFP/TeNT* mouse at postnatal day 21 colocalizes with anti-CaBP5 labelling in ON but not OFF bipolar cells[15]. **c–h,** Overview (**c–e**) and enlargements of ON sublamina (**f–h**) from postnatal day 21 *mGluR6-YFP/TeNT* (**c, d, f, g**) and wild-type (WT) retinas (**e, h**). YFP not shown in **d** and **g. i–k,** Representative traces (**i,** ON RGCs) and population data (mean + s.e.m., **j, k**) for sEPSCs recorded from RGCs at postnatal day 21. **l, m,** Spatial profile (**l,** 100 ms before spike) and time course (**m,** average of ten pixels with highest s.d.) of STAs from representative RGCs. **n, o,** Summary data (mean + s.e.m.) of receptive field extent (**n,** number of pixels with an s.d. that exceeded that of background pixels more than fourfold) and sensitivity (**o,** peak of average STA time course exemplified in **m**). Asterisk indicates $P < 0.01$.

(see Supplementary Information). Neurotransmission, however, is crucial for normal synaptic connectivity with the selected inputs.

In contrast to monostratified RGCs, bistratified RGCs maintain two narrowly stratified arborizations at distinct depths of the inner plexiform layer (Fig. 2k). The inner arbor is contacted by ON bipolar cells, the outer by OFF bipolar cells[3,4]. In wild-type mice, synaptic densities across both arborizations are correlated ($r = 0.86$; $P < 0.01$, $n = 8$ cells), presumably to balance ON and OFF light responses for each cell. If this balance was achieved by matching ON and OFF input

activity, one would predict that a lower density of synapses on the ON arbor of bistratified RGCs in *mGluR6-YFP/TeNT* mice would be accompanied by a decrease in synapses on its OFF arbor. Alternatively, if ON and OFF bipolar cells engaged in synaptic competition, one might expect the density of OFF arbor synapses to increase[1,2]. Instead, we found that synaptic densities on the ON dendrites of bistratified RGCs were reduced ($P < 0.04$) without changes to OFF dendrites ($P > 0.4$, Fig. 2k–p). These results show that synaptic connections on the two arborizations are established independently of each other and are sensitive to transmitter release locally, but not to differences between inputs onto separate arbors of the same cell.

Several presynaptic changes could accompany the reduced density of synapses on RGC dendrites in *mGluR6-YFP/TeNT* mice: (1) there could be fewer ON bipolar cells; (2) their axonal arborizations could be smaller and less complex; and (3) the density of synapses along axonal arborizations could be reduced. Comparison with *mGluR6-GFP* mice[14] showed that the density of ON bipolar cells in *mGluR6-YFP/TeNT* mice was unchanged (*mGluR6-YFP/TeNT*, 32,045 ± 1,173 cells mm^{-2}; *mGluR6-GFP*, 30,820 ± 610 cells mm^{-2}; $P > 0.3$, Supplementary Fig. 5). To assess the structure of single bipolar cell axons, we generated *mGluR6-tdTomato* mice, selected founders in which isolated ON bipolar cells were labelled and crossed them to *mGluR6-YFP/TeNT* mice. Most of the tdTomato-expressing bipolar cells were type 6 and type 7 ON cone bipolar cells or rod bipolar cells (Supplementary Fig. 6)[15]. Axonal arborizations of these cell types retained their characteristic stratification and branching patterns in *mGluR6-YFP/TeNT* compared to wild-type background (Fig. 3a–f). However, when we labelled presynaptic sites in bipolar cell axons with an antibody against C-terminal binding protein 2 (CtBP2), we found that their density in *mGluR6-YFP/TeNT* mice was reduced (Fig. 3g and h).

To determine whether the remaining synapses were structurally normal, we obtained electron micrographs from retinas of *mGluR6-YFP/TeNT* mice and wild-type littermates. Vesicles at bipolar cell active zones are tethered to synaptic ribbons, a common feature of neurons using graded voltage signals rather than action potentials[16]. Each bipolar cell synapse normally contains a single ribbon, but the signals that regulate ribbon localization and number remain unknown[16]. As expected, all bipolar cell active zones (25 of 25) in wild-type mice contained single ribbons (Fig. 3i, j and Supplementary Figs 7 and 8). By contrast, at about half of the ON bipolar cell active zones (12 of 25) in *mGluR6-YFP/TeNT* mice, we found more than one and up to four ribbons (Fig. 3i, j). This phenotype was restricted to ON bipolar cells; only single ribbons (11 of 11) were found in the OFF sublamina of *mGluR6-YFP/TeNT* mice (Supplementary Fig. 7). Because active zones and ribbons assemble by insertion of Piccolo–Bassoon transport vesicles and/or the arrival of precursor spheres, both of which do not contain VAMP2[17,18], it seems unlikely that TeNT interferes directly with this process. Instead, we propose that transmitter release normally elicits a signal that prohibits the placement of further ribbons at this active zone (see Supplementary Information).

Because the number of bipolar cells, their axons and the RGC dendrites they contact were normal in TeNT retinas, it remained unclear which developmental mechanisms account for the lower density of bipolar cell–RGC synapses observed at maturity (Fig. 2). To begin to address this question, we measured the number and density of excitatory synapses on ON RGC dendrites at different times during development (Fig. 4a, b). At postnatal day 7, when synaptic transmission between bipolar cells and RGCs is first observed[19], synapse numbers and densities were indistinguishable between wild-type and *mGluR6-YFP/TeNT* mice ($P > 0.4$) and gradually diverged over the following days towards the differences observed at maturity (Fig. 4a, b). To assess the dynamics that underlie this divergent development, we carried out live imaging experiments of RGCs biolistically labelled with tdTomato and PSD95–CFP in *mGluR6-YFP/TeNT* mice and wild-type littermates at postnatal day 9 (refs 20, 21). Time-lapse imaging revealed frequent synapse
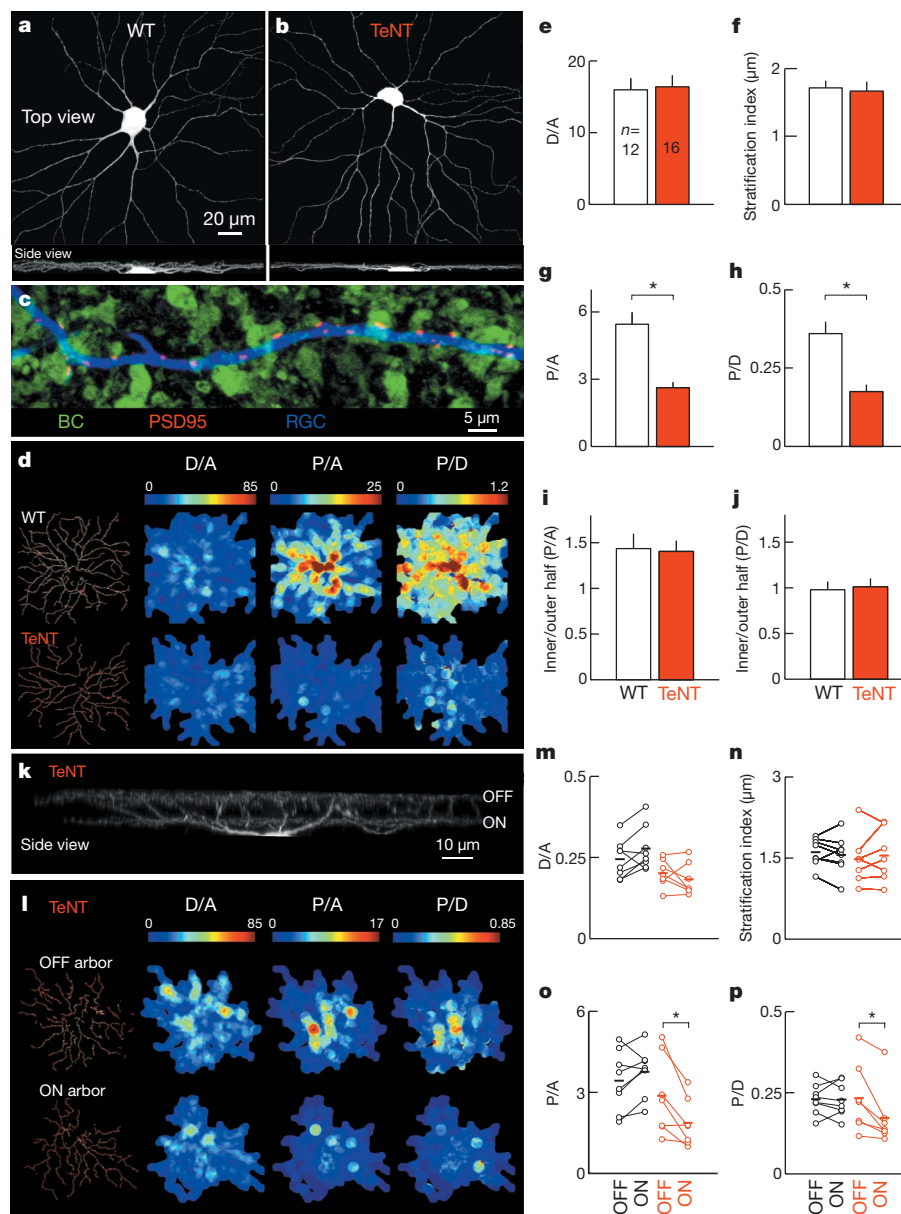
Figure 2 | Silencing ON bipolar cells reduces synapse number on RGC dendrites in an input-specific manner without changes to laminar targeting or branching. a, b, Large ON RGCs (tdTomato, postnatal day 21). **c,** RGC dendrite (tdTomato) in a TeNT retina (postnatal day 21) illustrating the localization of PSD95–CFP puncta to bipolar cell (BC, YFP) contacts (see also Supplementary Movie 1 and Supplementary Fig. 4). **d,** Maps of dendritic and glutamatergic postsynaptic densities for representative postnatal day 21 ON RGCs. **e–j,** Mean (±s.e.m.) RGC dendritic density (**e,** D/A) and stratification (**f,** see also Supplementary Fig. 3), areal (**g,** P/A) and linear (**h,** P/D) density of PSD95–CFP puncta, and centroperipheral gradients (**i** and **j**) of PSD95–CFP puncta across dendritic fields for large ON RGCs[13].

k, Bistratified ON-OFF RGC (tdTomato, postnatal day 21) in TeNT retina. **l,** Maps of dendritic and glutamatergic postsynaptic densities for OFF and ON arborization for a representative bistratified RGC in an *mGluR6-YFP/ TeNT* mouse. **m–p,** Comparison of dendritic density (**m,** D/A) and stratification (**n**), and areal (**o,** P/A) and linear (**p,** P/D) density of PSD95–CFP puncta between the OFF and ON arbor of bistratified RGCs from wild-type (black) and *mGluR6-YFP/TeNT* (red) mice . Mean indicated by bold horizontal lines. Asterisk indicates *P* < 0.01. D/A, Dendritic length (μm) per 100 μm²; P/A, puncta per 100 μm²; P/D, puncta per dendritic length (μm).

formation (Fig. 4c–e and Supplementary Fig. 9) and elimination (Fig. 4f–h) events. The synaptic turnover rate was 27 ± 5.8% per day with a net increase in synapses of 11 ± 6.5% per day in wild-type mice. Surprisingly, in *mGluR6-YFP/TeNT* mice the rate of synapse formation was reduced several fold (Fig. 4i), whereas synapse elimination was indistinguishable from wild-type littermates (Fig. 4j). We verified that >90% of newly formed synapses were present in more than one time-lapse image to make sure that the reduced rate of synapse formation we report was not caused by a shortening of synapse lifetimes below our sampling interval. Moreover, the difference in the rate of synapse formation between TeNT and wild-type retinas (36 ± 13%; TeNT/wild type in per cent, Fig. 4i) was similar to the

difference in synapse density increase estimated by linear regression from postnatal day 7 to 21 (23 ± 13%; TeNT/wild type in per cent, Fig. 4a). Together these results show that inhibition of glutamate release from TeNT-expressing ON bipolar cells selectively lowers the rate at which new synapses are established during development, thus accounting for the reduced density of bipolar cell–RGC synapses observed in mature *mGluR6-YFP/TeNT* mice.

The self-organization of neurons into complex circuits during development is thought to depend in part on neurotransmission. The most prominent model of activity-dependent circuit development proposes that converging inputs engage in activity-mediated competitions, the outcome of which is determined by differences in
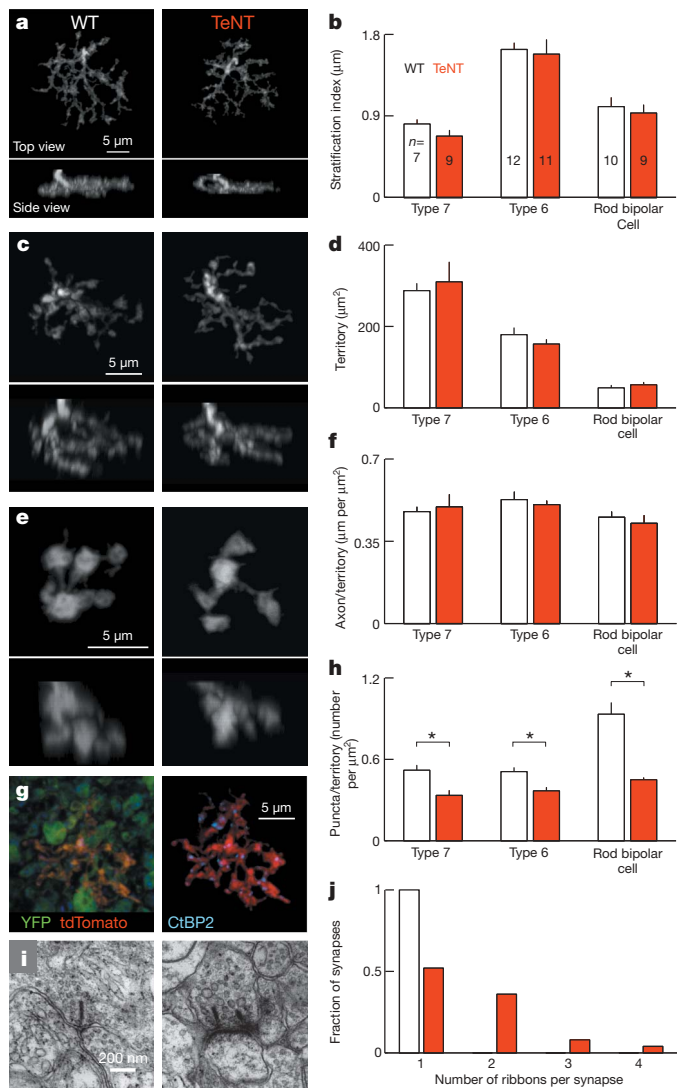
Figure 3 | Axonal morphology is normal, but multiple ribbons accumulate at fewer synapses in TeNT-expressing ON bipolar cells. a–f, Axon arborizations of representative type 7 (**a**) and type 6 (**c**) ON cone bipolar cells, and rod bipolar cells (**e**) labelled in an *mGluR6-tdTomato* mouse in wild-type (left) or *mGluR6-YFP/TeNT* (right) background at postnatal day 21. TdTomato signal was used to mask arborizations of single bipolar cells. Summary data (mean + s.e.m.) comparing stratification (**b**), lateral territories (**d**) and axonal density (**f**) between wild-type and *mGluR6-YFP/TeNT* background. **g**, Synapses on bipolar cell terminals labelled with anti-CtBP2. TdTomato signal shown in *mGluR6-YFP/TeNT* background (left) was used as a mask to isolate ribbon synapses of an individual axonal arbor (right). **h**, Mean (+s.e.m.) density of synapses for the different bipolar cell types. Asterisk indicates $P < 0.01$. **i**, Electron micrographs of ON bipolar cell ribbon synapses of wild-type (right) and *mGluR6-YFP/TeNT* (left) littermates at postnatal day 21. **j**, Histogram of the number of ribbons per synapse for wild-type (white, $n = 25$) and *mGluR6-YFP/TeNT* (red, $n = 25$) mice.

synapse elimination[1,2]. This is supported by studies of the neuromuscular junction. During development, each neuromuscular junction receives convergent innervation from multiple axons of which all but one are eventually retracted. When transmitter release from a subset of these axons is inhibited, the outcome of competition is biased towards removal of silenced axons[22,23]. In contrast, studies of RGC axons in the zebrafish tectum showed that activity could regulate branch addition rather than retraction[24], suggesting that different rules might apply in different circuits. In this study, we modified synaptic activity directly while imaging synaptic dynamics in the developing retina. In doing so we find that during the assembly of
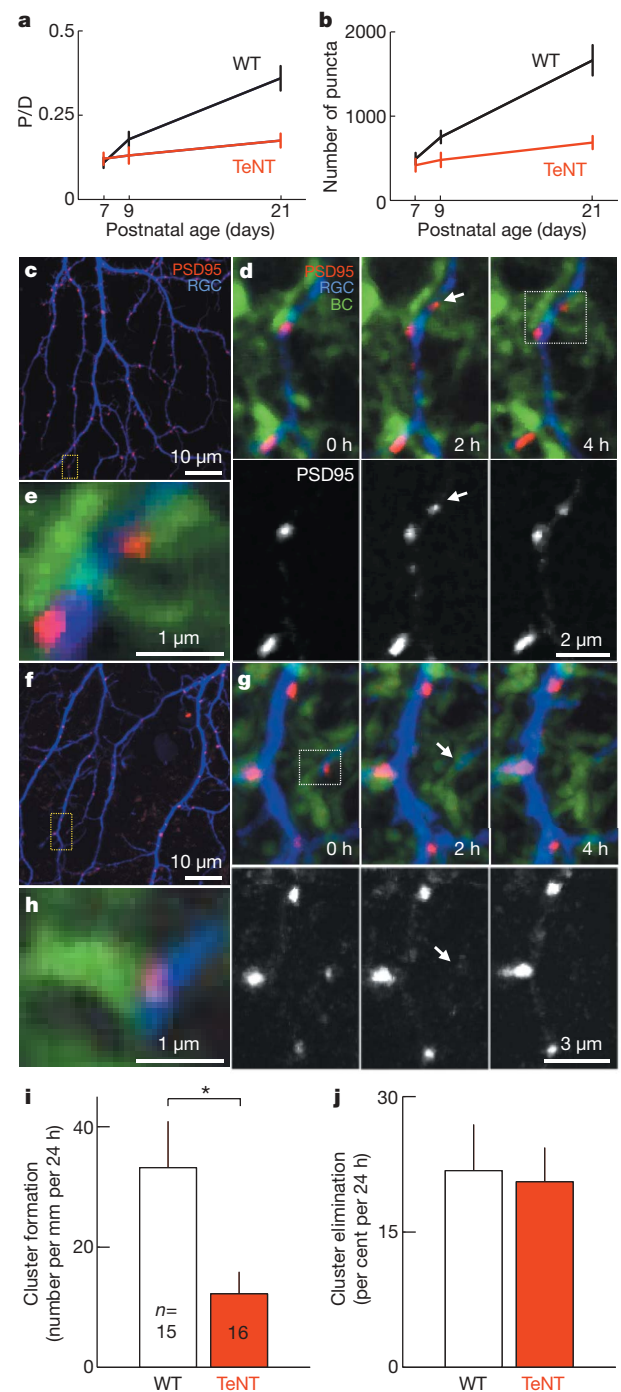
Figure 4 | Transmitter release regulates synapse formation but not elimination, causing a gradual divergence of synaptic development between wild-type and *mGluR6-YFP/TeNT* mice. a, b, Developmental increase in density (**a**) and number (**b**) of PSD95–CFP puncta on ON RGCs (5–16 cells per time point and genotype). **c**, First image of a time-lapse series of a postnatal day 9 RGC labelled with tdTomato (blue) and glutamatergic postsynaptic densities labelled with PSD95–CFP (red). **d**, Time series of the region indicated by a yellow box in **c** in which a PSD95 cluster forms (arrow). YFP expression in bipolar cells is shown in green. The PSD95–CFP channel is shown in the lower panels. **e**, Higher magnification of a single plane of the region indicated by the white box in **d** showing that the new punctum formed at an apposition of bipolar cell axon and RGC dendrite. **f–h**, analogous to **c–e** for loss of a PSD95 cluster. **i, j**, Summary data (mean ± s.e.m.) for the rate of PSD95–CFP cluster formation (**i**) and elimination (**j**). Asterisk indicates $P < 0.05$.

retinal circuits, glutamate release regulates the formation but not elimination of synapses between bipolar cells and RGCs. In addition, we find that rather than engaging in competition, the axons of ON

and OFF bipolar cells refine their connectivity with RGCs independent of one another, even when they converge stably onto a single postsynaptic cell. Finally, neurotransmission seems to selectively affect synaptic but not axonal or dendritic, development in this system. Which factors determine the different rules that seem to guide activity-dependent development of different neural circuits? Part of the answer might be found in the distinct architectures of early neural circuits that are set up before neurotransmission. Accordingly, the early spatial separation of ON and OFF bipolar cell axons in the inner plexiform layer might serve to constrain the role of activity in synaptic remodelling. It is tempting to speculate that laminar circuit architecture throughout the nervous system[25] might similarly serve to limit activity-dependent competition between parallel circuits that process distinct information, while allowing activity to fine-tune wiring within each information channel.

METHODS SUMMARY

Transgenic mice. We generated two transgenic mouse lines. In one (*mGluR6-YFP/TeNT*) a ~9-kb fragment of the *mGluR6* promoter was used to express yellow fluorescent protein (YFP) and the light chain of tetanus toxin (TeNT) selectively in retinal ON bipolar cells. In the other (*mGluR6-tdTomato*) the ~9-kb *mGluR6* promoter fragment drove expression of the red fluorescent protein tandem dimer Tomato (tdTomato).

Immunohistochemistry. Paraformaldehyde (4%)-fixed retinal flat-mount preparations or vibratome slices were labelled by standard immunohistochemistry procedures.

Electrophysiology. We performed whole-cell patch-clamp recordings from RGCs in retinal flat-mount preparations to analyse spontaneous excitatory synaptic transmission and used multi-electrode array recordings to characterize light responses during white noise presentation from an organic light-emitting display (OLED).

Imaging and analysis. Images of fixed tissue were acquired on an Olympus FV1000 confocal microscope using a 1.35 NA ×60 oil objective. Time-lapse imaging experiments on retinas of *mGluR6-YFP/TeNT* mice and wild-type littermates were performed in parallel on an Olympus FV1000 confocal microscope (*mGluR6-YFP/TeNT*) and either an Olympus FV300 confocal microscope or a custom-built two-photon microscope (wild type) using 1.1 NA ×60 water objectives. Image stacks were analysed using Amira (Mercury Computer System), Imaris (Bitplane) and custom software written in MATLAB (The Mathworks).

Electron microscopy. Retinas were fixed using standard electron microscopy fixation techniques, cut into pieces and en bloc-stained in uranyl acetate. Thin sections were cut and stained with Reynolds lead citrate.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Wong, R. O. L. & Lichtman, J. W. in *Fundamental Neuroscience* 2nd edn (eds Squire, L. R. *et al.*) Ch. 20, 533–554 (Academic Press, 2002).
2. Miller, K. D. Synaptic economics: competition and cooperation in synaptic plasticity. *Neuron* 17, 371–374 (1996).
3. Masland, R. H. The fundamental plan of the retina. *Nature Neurosci.* 4, 877–886 (2001).
4. Wassle, H. Parallel processing in the mammalian retina. *Nature Rev. Neurosci.* 5, 747–757 (2004).
5. Ueda, Y., Iwakabe, H., Masu, M., Suzuki, M. & Nakanishi, S. The mGluR6 5′ upstream transgene sequence directs a cell-specific and developmentally regulated expression in retinal rod and ON-type cone bipolar cells. *J. Neurosci.* 17, 3014–3023 (1997).
6. Schiavo, G. *et al.* Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* 359, 832–835 (1992).
7. Harms, K. J. & Craig, A. M. Synapse composition and organization following chronic activity blockade in cultured hippocampal neurons. *J. Comp. Neurol.* 490, 72–84 (2005).
8. Chichilnisky, E. J. A simple white noise analysis of neuronal light responses. *Network* 12, 199–213 (2001).
9. Maslim, J., Webster, M. & Stone, J. Stages in the structural differentiation of retinal ganglion cells. *J. Comp. Neurol.* 254, 382–402 (1986).
10. Coombs, J. L., Van Der List, D. & Chalupa, L. M. Morphological properties of mouse retinal ganglion cells during postnatal development. *J. Comp. Neurol.* 503, 803–814 (2007).
11. Wang, G. Y., Liets, L. C. & Chalupa, L. M. Unique functional properties of On and Off pathways in the developing mammalian retina. *J. Neurosci.* 21, 4310–4317 (2001).
12. Nelson, R., Famiglietti, E. V. Jr & Kolb, H. Intracellular staining reveals different levels of stratification for on- and off-center ganglion cells in cat retina. *J. Neurophysiol.* 41, 472–483 (1978).
13. Morgan, J. L., Schubert, T. & Wong, R. O. L. Developmental patterning of glutamatergic synapses onto retinal ganglion cells. *Neural Develop.* 3, 8 (2008).
14. Morgan, J. L., Dhingra, A., Vardi, N. & Wong, R. O. L. Axons and dendrites originate from neuroepithelial-like processes of retinal bipolar cells. *Nature Neurosci.* 9, 85–92 (2006).
15. Ghosh, K. K., Bujan, S., Haverkamp, S., Feigenspan, A. & Wassle, H. Types of bipolar cells in the mouse retina. *J. Comp. Neurol.* 469, 70–82 (2004).
16. Sterling, P. & Matthews, G. Structure and function of ribbon synapses. *Trends Neurosci.* 28, 20–29 (2005).
17. Regus-Leidig, H., Tom Dieck, S., Specht, D., Meyer, L. & Brandstatter, J. H. Early steps in the assembly of photoreceptor ribbon synapses in the mouse retina: the involvement of precursor spheres. *J. Comp. Neurol.* 512, 814–824 (2009).
18. Zhai, R. G. *et al.* Assembling the presynaptic active zone: a characterization of an active one precursor vesicle. *Neuron* 29, 131–143 (2001).
19. Johnson, J. *et al.* Vesicular neurotransmitter transporter expression in developing postnatal rodent retina: GABA and glycine precede glutamate. *J. Neurosci.* 23, 518–529 (2003).
20. Okabe, S., Miwa, A. & Okado, H. Spine formation and correlated assembly of presynaptic and postsynaptic molecules. *J. Neurosci.* 21, 6105–6114 (2001).
21. Niell, C. M., Meyer, M. P. & Smith, S. J. *In vivo* imaging of synapse formation on a growing dendritic arbor. *Nature Neurosci.* 7, 254–260 (2004).
22. Buffelli, M. *et al.* Genetic evidence that relative synaptic efficacy biases the outcome of synaptic competition. *Nature* 424, 430–434 (2003).
23. Kasthuri, N. & Lichtman, J. W. The role of neuronal identity in synaptic competition. *Nature* 424, 426–430 (2003).
24. Hua, J. Y., Smear, M. C., Baier, H. & Smith, S. J. Regulation of axon growth *in vivo* by activity-based competition. *Nature* 434, 1022–1026 (2005).
25. Sanes, J. R. & Yamagata, M. Formation of lamina-specific synaptic connections. *Curr. Opin. Neurobiol.* 9, 79–87 (1999).

METHODS

Transgenic mice. The following transgenic constructs were assembled using standard cloning techniques. *mGluR6-loxP-YFPstop-loxP-TeNT-CFP*: a 9.5-kb fragment of the *mGluR6* promoter[5] (gift of S. Nakanishi) was placed upstream of a 2.5-kb *YFPstop* cassette which was flanked by two *loxP* sites. The *YFPstop* cassette contained the coding sequence of enhanced yellow fluorescent protein followed by the 3′ UTR and transcription stop site of the rabbit β-globin gene. A 2.3-kb *TeNT–CFP* cassette encoding a carboxy-terminal fusion of enhanced cyan fluorescent protein to the light chain of tetanus toxin (TeNT–CFP, gift of R.W. Burgess and J.R. Sanes) was inserted 3′ of the second *loxP* site. *mGluR6-tdTomato*: the 9.5-kb *mGluR6* promoter fragment was positioned upstream of the coding sequence of the red fluorescent protein tdTomato[26]. Linearized constructs were injected into pronuclei of murine oocytes to generate transgenic mouse lines. We refer to these lines as *mGluR6-YFP/TeNT* and *mGluR6-tdTomato* throughout the text. PCR genotyping was performed using primers to enhanced yellow fluorescent protein (5′-GACTTCTTCAAGTCCGCCATG CC-3′ and 5′-GTGATCCCGGCGGCGGTCACG-3′) and tdTomato (5′-CTCC TCCGAGGACAACAACATGG and 5′-CTTGGTCACCTTCAGCTTGGCGG-3′).

We initially crossed *mGluR6-YFP/TeNT* mice to *Pax6_α-cre* mice[27] to excise the *YFPstop* cassette and allow expression of TeNT–CFP. Although retinal ON bipolar cells in mice carrying both transgenes showed varying mixtures of YFP and CFP fluorescence, the TeNT substrate synaptobrevin/VAMP2 was cleaved uniformly in these cells (data not shown). To our surprise, this was also true for ON bipolar cells in the dorso-ventral wedge of the retina which in *Pax6_α-cre* mice do not express Cre recombinase[27]. We next verified that VAMP2 was cleaved even in ON bipolar cells of *mGluR6-YFP/TeNT* mice that had not been crossed to *Pax6_α-cre* mice (Fig. 1c–h). In addition, we could detect dim CFP fluorescence in these cells (data not shown). Weak expression of TeNT–CFP in these mice could be caused by occasional transcription past the *YFPstop* cassette or spontaneous recombination and excision of this cassette in one or a few of the copies of the transgene array. Because weak TeNT–CFP expression was sufficient to block release from ON bipolar cells (Fig. 1i, j), we used *mGluR6-YFP/TeNT* mice without crossing to *Pax6_α-cre* mice throughout this study. We observed some variation in the number of bipolar cells expressing the transgene even among littermates. In particular, some mice showed patchy expression patterns across the retina that were most likely due to transgene-induced gene silencing. For the present study we excluded those retinas. To do so we monitored YFP expression pattern in both electrophysiological and imaging experiments.

Tissue preparation. All experiments were carried out following the guidelines of the Institutional Animal Care and Use Committee at the University of Washington and in compliance with the National Institute of Health Guide for the Care and Use of Laboratory Animals. Mice, deeply anaesthetized with isoflurane, were decapitated and enucleated. Each cornea was punctured with a 30 gauge needle and the eyes placed in cold oxygenated mouse artificial cerebrospinal fluid (mACSF) containing 119 mM NaCl, 2.5 mM KCl, 2.5 mM CaCl$_2$, 1.3 mM MgCl$_2$, 1 mM NaH$_2$PO$_4$, 11 mM glucose and 20 mM HEPES. mACSF was brought to pH 7.35 with NaOH. For vibratome sections, the lens and vitreous were removed and the remaining eyecup fixed for 30 min in 4% paraformaldehyde in mACSF. For electron microscopy, eyecups were placed in 2% glutaraldehyde and 2% paraformaldehyde in 0.1 M sodium cacodylate (pH 7.4) for 10 min. For flat-mount preparation, the retina was isolated and mounted ganglion cell side up on membrane discs (Millipore) before fixation (30 min in 4% paraformaldehyde in mACSF), biolistic transfection or patch-clamp electrophysiology.

Immunohistochemistry. Fixed retinal flat mounts or vibratome slices were labelled with the antibodies anti-CaBP5 (1:1,000, gift of F. Haessler and K. Palczewski), anti-VAMP2 (1:1,000, Synaptic Systems), anti-CtBP2 (1:1,000, BD Bioscience Pharmingen) and anti-GluR2/3 (1:1,000, Upstate). Secondary antibodies were Alexa 568 or Alexa 633 conjugates (1:1,000, Invitrogen).

Biolistic transfection. Gold particles (12.5 mg, 1.0 or 1.6 µm diameter, Bio-Rad) were coated with cytomegalovirus-promoter-driven PSD95–CFP (10 µg) and tdTomato (20 µg) plasmids and delivered to RGCs in a flat-mount preparation using a Helios gene gun (~40 psi, Bio-Rad). Transfected retinas were incubated in mACSF in a humid oxygenated chamber at ~33 °C for 18–24 h, and transferred to a live imaging chamber or fixed for 30 min in 4% paraformaldehyde in mACSF.

Confocal imaging of fixed tissue and analysis. RGC image stacks were acquired on an Olympus FV1000 laser scanning confocal microscope using a 1.35 NA ×60 oil objective at a voxel size of 0.103 × 0.103 × 0.3 µm (*x-y-z*) except for the data set in Supplementary Fig. 3 which was imaged using a 0.8 NA ×25 oil objective at a voxel size of 0.247 × 0.247 × 0.5 µm (*x-y-z*). Voxel size for image stacks of bipolar cell axons acquired using a 1.35 NA ×60 oil objective was 0.041 × 0.041 × 0.3 µm (*x-y-z*).

Morphometric analysis of RGC dendrites and mapping of excitatory synapses were carried out largely as described previously[13]. Dendritic territories were defined by dilating the dendritic skeleton with a 10 µm diameter disk. Within the dendritic territory of each cell, dendritic length (µm) per 100 µm^2 (D/A), puncta per 100 µm^2 (P/A) and puncta per dendritic length (µm; P/D) were measured. The stratification index was defined differently from the previous study[13], as the s.d. of *z*-positions of the ~0.5 µm long segments that form the dendritic skeleton. To reduce the influence of irregularities in the tissue, the average *z*-position used in the calculation of the s.d. was updated for each segment as the average *z*-position of segments within 30 µm in *x-y* of the current segment. In order to compare distributions only across the stratified arborizations, the analysis excluded dendrites within 10 µm from the cell body.

To isolate axonal arborizations of single bipolar cells, *mGluR6-YFP/TeNT* mice were crossed to a line of *mGluR6-tdTomato* mice in which ON bipolar cells were labelled at a low density. Presynaptic release sites were labelled with antibodies against the ribbon protein CtBP2. The tdTomato signal was used to mask (Amira, Mercury Computer Systems) parts of the image stack representing a single bipolar cell. To measure arbor length, axons were skeletonized based on the binary mask of the tdTomato signal using an algorithm similar to the one used for RGC dendrites[13]. Arbor stratification was measured as above. Axonal territories were defined by the smallest convex polygon encompassing the arbor in a *z*-projection. To automate counting of synapses on bipolar cell axons, the masked CtBP2 image was median-filtered and thresholded at every other grey value from 255 to 0. For each threshold, objects containing <6 or >343 connected voxels were removed. The position of the peak intensity of each remaining object was noted in a binary peak map. Within the iterative thresholding loop a summary image was accumulated in which the intensity of each voxel signified the number of times it passed threshold when associated with an object of acceptable size (6–343 voxels). In the next step, objects of connected voxels in this summary image were identified and split into multiple objects if they were associated with more than one entry in the peak map. These objects will be referred to as potential synapses. The voxels associated with a potential synapse were defined as all voxels of the respective object whose intensity exceeded half the intensity of its peak. For each potential synapse two parameters were calculated: strength and surprise. Strength was defined as the summed intensity across all voxels of one synapse in the summary image divided by their average brightness in the median-filtered CtBP2 image. To calculate surprise, we first determined the expected intensity of the CtBP2 channel given the value of the tdTomato channel from linear regression of both channels at voxels outside potential synapses. Next the difference between the observed brightness in the CtBP2 channel and its expected brightness for each voxel of a potential synapse was calculated and normalized by the expected brightness. The surprise of a potential synapse was defined as the average of this normalized difference for the brightest half of its voxels. For each bipolar cell, all potential synapses were depicted in a coordinate system of strength and surprise, and the cluster of acceptable synapses selected via a graphical user interface. For each cell this selection was verified by overlaying the position of acceptable synapses on the CtBP2 image. We confirmed that this procedure correctly identified >90% of the synaptic puncta identified by eye on a randomly chosen subset of cells. All scripts for customized image analysis were written in MATLAB (The MathWorks).

Confocal and two-photon live imaging and analysis. Time-lapse imaging experiments on retinas of *mGluR6-YFP/TeNT* mice and wild-type littermates (~postnatal day 9) were performed using an Olympus FV1000 confocal microscope (*mGluR6-YFP/TeNT*, three-colour images) and either an Olympus FV300 confocal or a custom-built two-photon microscope (wild type, two-colour images) using 1.1 NA ×60 water objectives. Flat-mounted retinas were continuously perfused with oxygenated mACSF and heated to 31–33 °C. Image stacks were acquired every 2 h for up to 14 h. Voxel size was between 0.103 × 0.103 × 0.3 µm (*x-y-z*) and 0.115 × 0.115 × 0.3 µm (*x-y-z*). For analysis image stacks were median-filtered to reduce photomultiplier noise. PSD95–CFP puncta were identified in image stacks by eye and manually tracked through time series. The synaptic turnover rate was calculated as ($N_{formations} + N_{eliminations}$)/($2 \times N_{total}$), where $N_{formations}$ is the number of new synapses, $N_{eliminations}$ is the number of lost synapses and N_{total} is the sum of all synapses. Dendrites were reconstructed and their length measured using Imaris (Bitplane). To avoid biasing measurement of synapse formation and stability by differences in the synapse number at the beginning of the imaging experiment, the rate of formation was normalized by the length of the dendrite and the rate of synapse elimination was expressed as the percentage of puncta that were lost.

Electron microscopy. After the initial fixation (described under Tissue preparation) retinas were cut into ~1 mm^2 pieces and left in fixative overnight at 4 °C. The pieces of retina were then washed four times for 5 min in sodium cacodylate buffer and post-fixed for 45 min in buffered 1% osmium tetroxide. Next, the

tissue was washed four times for 5 min in dH$_2$O, en bloc-stained in 1% aqueous uranyl acetate, and washed twice again in dH$_2$O before being dehydrated through a series of ethanol and propylene oxide and embedded in Epon Araldite resin. Thin sections (70 nm) were cut, placed on Formvar-coated slot grids and stained with Reynolds lead citrate.

Patch-clamp recording. Retinal flat mounts were prepared as described above and bathed in bicarbonate-buffered mACSF containing 125 mM NaCl, 2.5 mM KCl, 2 mM CaCl$_2$, 1 mM MgCl$_2$, 1.25 mM NaH$_2$PO$_4$, 11 mM glucose and 26 mM NaHCO$_3$ (equilibrated with 95% O$_2$ and 5% CO$_2$). Whole-cell recordings were performed with electrodes (4–8 MΩ) filled with 120 mM Cs-gluconate, 1 mM CaCl$_2$, 1 mM MgCl$_2$, 10 mM Na-HEPES, 11 mM EGTA, 10 mM TEA-Cl and 2 mM QX314 (pH 7.2 adjusted with CsOH). Data were acquired using an Axopatch 200B amplifier (Molecular Devices), low-pass filtered at 2 kHz and digitized at 5 kHz. sEPSC were recorded at –60 mV, the reversal potential of chloride currents in our recording conditions. Area and amplitude thresholds (Mini Analysis, Synaptosoft) were optimized to detect >90% of the events identified by eye in excerpts from each recording. For overlapping events, the baseline for amplitude measurement of each event was estimated from exponential decay extrapolation of the previous event. RGCs were classified as ON or OFF based on their stratification pattern that was visualized by including sulforhodamine B (0.005%) in the recording pipette.

Multi-electrode array recording and visual stimulation. Retinas were isolated as described above, transferred to a chamber containing the array (HD MEA, Multi Channel Systems), floated onto the electrodes RGC-side down and held in place by a transparent tissue culture membrane (Corning) and a platinum ring. During recordings the tissue was perfused with bicarbonate-buffered mACSF equilibrated with 95% O$_2$ and 5% CO$_2$. The temperature of the bath was maintained at 31–33 °C. Signals were band-pass filtered between 0.1 and 3 kHz and digitized at 10 kHz. Spikes were detected by negative thresholds and sorted into trains representing the activity of single cells (Offline Sorter, Plexon). Only cells for which <1% of interspike intervals were <2 ms were used in our analysis. Duplicates, that is, cells recorded on multiple electrodes were identified by cross-correlation of spike trains, and only the train with the highest number of spikes was kept. Stimuli composed of a lattice of squares, each flickering randomly and independently at 25 Hz, were presented on a monochrome yellow organic light-emitting diode microdisplay (OLED, eMagin) at a mean equivalent intensity of ~10^4 photo-isomerizations per middle-wavelength-sensitive cone per second. Each square had a side length of 50 μm on the retina. Response properties of RGCs were characterized by the spike-triggered average (STA) stimulus calculated from reverse correlation of its spike train to the stimulus[8,28].

Statistics. We used either Wilcoxon–Mann–Whitney rank sum or, in case of paired samples, Wilcoxon signed-rank tests to assess statistical significance of differences between groups.

26. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnol.* **22**, 1567–1572 (2004).
27. Marquardt, T. *et al.* Pax6 is required for the multipotent state of retinal progenitor cells. *Cell* **105**, 43–55 (2001).
28. Kerschensteiner, D. *et al.* Genetic control of circuit function: *Vsx1* and *Irx5* transcription factors regulate contrast adaptation in the mouse retina. *J. Neurosci.* **28**, 2342–2352 (2008).

LETTERS

In vitro and *in vivo* characterization of new swine-origin H1N1 influenza viruses

Yasushi Itoh[1], Kyoko Shinya[2], Maki Kiso[3], Tokiko Watanabe[4], Yoshihiro Sakoda[5], Masato Hatta[4], Yukiko Muramoto[6], Daisuke Tamura[3], Yuko Sakai-Tagawa[3], Takeshi Noda[7], Saori Sakabe[3], Masaki Imai[4], Yasuko Hatta[4], Shinji Watanabe[4], Chengjun Li[4], Shinya Yamada[3], Ken Fujii[6], Shin Murakami[3], Hirotaka Imai[3], Satoshi Kakugawa[3], Mutsumi Ito[3], Ryo Takano[3], Kiyoko Iwatsuki-Horimoto[3], Masayuki Shimojima[3], Taisuke Horimoto[3], Hideo Goto[3], Kei Takahashi[3], Akiko Makino[2], Hirohito Ishigaki[1], Misako Nakayama[1], Masatoshi Okamatsu[5], Kazuo Takahashi[8], David Warshauer[9], Peter A. Shult[9], Reiko Saito[10], Hiroshi Suzuki[10], Yousuke Furuta[11], Makoto Yamashita[12], Keiko Mitamura[13], Kunio Nakano[13], Morio Nakamura[13], Rebecca Brockman-Schneider[14], Hiroshi Mitamura[15], Masahiko Yamazaki[16], Norio Sugaya[17], M. Suresh[4], Makoto Ozawa[4,7], Gabriele Neumann[4], James Gern[14], Hiroshi Kida[5], Kazumasa Ogasawara[1] & Yoshihiro Kawaoka[2,3,4,6,7,18]

Influenza A viruses cause recurrent outbreaks at local or global scale with potentially severe consequences for human health and the global economy. Recently, a new strain of influenza A virus was detected that causes disease in and transmits among humans, probably owing to little or no pre-existing immunity to the new strain. On 11 June 2009 the World Health Organization declared that the infections caused by the new strain had reached pandemic proportion. Characterized as an influenza A virus of the H1N1 subtype, the genomic segments of the new strain were most closely related to swine viruses[1]. Most human infections with swine-origin H1N1 influenza viruses (S-OIVs) seem to be mild; however, a substantial number of hospitalized individuals do not have underlying health issues, attesting to the pathogenic potential of S-OIVs. To achieve a better assessment of the risk posed by the new virus, we characterized one of the first US S-OIV isolates, A/California/04/09 (H1N1; hereafter referred to as CA04), as well as several other S-OIV isolates, *in vitro* and *in vivo*. In mice and ferrets, CA04 and other S-OIV isolates tested replicate more efficiently than a currently circulating human H1N1 virus. In addition, CA04 replicates efficiently in non-human primates, causes more severe pathological lesions in the lungs of infected mice, ferrets and non-human primates than a currently circulating human H1N1 virus, and transmits among ferrets. In specific-pathogen-free miniature pigs, CA04 replicates without clinical symptoms. The assessment of human sera from different age groups suggests that infection with human H1N1 viruses antigenically closely related to viruses circulating in 1918 confers neutralizing antibody activity to CA04. Finally, we show that CA04 is sensitive to approved and experimental antiviral drugs, suggesting that these compounds could function as a first line of defence against the recently declared S-OIV pandemic.

Sequence analyses of recently emerged swine-origin H1N1 viruses (S-OIVs) revealed the absence of markers associated with high pathogenicity in avian and/or mammalian species, such as a multibasic haemagglutinin (HA) cleavage site[2] or lysine at position 627 of the PB2 protein[3]. To characterize the new viruses *in vitro* and *in vivo*, we amplified the following S-OIVs in Madin–Darby canine kidney (MDCK) cells: A/California/04/09 (CA04), A/Wisconsin/WSLH049/ 09 (WSLH049), A/Wisconsin/WSLH34939/09 (WSLH34939), A/Netherlands/603/09 (Net603) and A/Osaka/164/09 (Osaka164). WSLH34939 was isolated from a patient who required hospitalization, whereas the remaining viruses were isolated from mild cases. These viruses represent the currently recognized neuraminidase (NA) variants among S-OIVs: CA04, NA-106V, NA-248N; Osaka164, NA-106I, NA-248N; WSLH049, NA-106I, NA-248D; WSLH34939, NA-106I, NA-248D; and Net603, NA-106V, NA-248N.

In MDCK cells and primary human airway epithelial cells, CA04 grew to titres comparable to those typically obtained for contemporary human H1N1 influenza viruses (Supplementary Fig. 1). Confocal, transmission electron and scanning electron microscopy revealed virions of remarkably filamentous shape (Supplementary Fig. 2), in marked contrast to the spherical shape observed with negatively stained virions (http://www.cdc.gov/h1n1flu/images.htm). The biological significance of the morphology of CA04 remains unknown.

To evaluate the pathogenicity of S-OIV in mammalian models, we conducted studies in mice, ferrets, non-human primates and pigs. BALB/c mice intranasally infected with a high dose ($>10^4$ plaque-forming units (p.f.u.)) of CA04 (Supplementary Fig. 3) experienced weight loss and those infected with the highest dose of this virus were humanely killed, in contrast to animals infected with a recent human H1N1 virus (A/Kawasaki/UTK-4/09, KUTK-4). The 50% mouse lethal dose (MLD_{50}) was $10^{5.8}$ p.f.u. for CA04 and $>10^{6.6}$ p.f.u. for

KUTK-4. For the additional S-OIV isolates tested, the MLD_{50} values were $>10^{6.4}$ p.f.u. for Osaka164, $>10^{6.6}$ p.f.u. for WSLH049, $10^{4.5}$ p.f.u. for WSLH34939 and $>10^{5.8}$ p.f.u. for Net603.

On day 3 after infection of mice, similar titres were detected in nasal turbinates of mice infected with 10^5 p.f.u. of S-OIVs or KUTK-4 (Supplementary Table 1); however, S-OIVs replicated more efficiently in the lungs of infected animals, which may account for the prominent bronchitis and alveolitis with viral antigen on day 3 after infection with CA04 (Supplementary Fig. 4a, b). On day 6 after infection, virus titres followed a similar trend and the lungs of CA04-infected mice showed bronchitis and alveolitis with viral antigen, although signs of regeneration were apparent (Supplementary Fig. 4c). We detected viral-antigen-positive bronchial epithelial cells, but not alveolar cells, on day 3 after infection of mice infected with KUTK-4 (Supplementary Fig. 4e). By day 6, infection in KUTK-4-inoculated mice had progressed to bronchitis and peribronchitis; however, viral antigen was rarely detected in these lesions (Supplementary Fig. 4f).

There were marked differences in the induction of pro-inflammatory cytokines in the lungs of mice infected with CA04 compared with KUTK-4 (Supplementary Fig. 5a–c). Infection with KUTK-4 resulted in limited induction of pro-inflammatory cytokines/chemokines in the lungs, in marked contrast to infection with CA04. Increased production of interleukin-10 (IL-10; Supplementary Fig. 5a) in lungs of CA04-infected mice at day 6 after infection probably reflects a host response to dampen over-exuberant pulmonary inflammation and promote tissue repair. Infection with CA04 led to strong induction of both interferon-γ (IFN-γ) and IL-4 in the lungs. The selective induction of the T_H2 cytokine IL-5 in CA04-infected, but not in KUTK-4-infected, mice on day 6 after infection is noteworthy (Supplementary Fig. 5b), but further studies are needed to understand the relevance of this finding to viral control. IL-17 has been reported to have a role in protection against lethal influenza and also in eliciting inflammatory responses[4,5]; however, the enhanced viral replication and lung pathology observed in CA04-infected mice was not linked to dysregulated IL-17 production.

Cynomolgus macaques (*Macaca fascicularis*) have been used to study highly pathogenic avian H5N1 viruses[6,7] and the 1918 pandemic virus[8]. Infection of cynomolgus macaques with CA04 (see Methods for detailed procedures) resulted in a more prominent increase in body temperature than infection with KUTK-4 (Supplementary Fig. 6). This difference might originate from the observed differences in virus titres (Table 1 and Supplementary Table 2). No remarkable difference in body weight loss was found between the two groups (data not shown). CA04 replicated efficiently in the lungs and other respiratory organs of infected animals, similar to highly pathogenic influenza viruses[6,8] (Table 1). By contrast, conventional human influenza viruses are typically limited in their replicative ability in the lungs

of infected primates[6,8] (Table 1), although a seasonal H1N1 virus was isolated from one animal on day 7 after infection. Pathological examination revealed that CA04 caused more severe lung lesions than did KUTK-4 (Fig. 1 and Supplementary Fig. 7). On day 3 after infection with CA04, alveolar spaces were occupied by oedematous exudate and inflammatory infiltrates (Fig. 1a, b); severe thickening of alveolar walls was also observed (Fig. 1b). Viral-antigen-positive cells were distributed in the inflammatory lesions, and many of these cells were elongated with thin cytoplasm and hemming around the alveolar wall, indicating type I pneumocytes (Fig. 1c). In addition to type I pneumocytes, CA04 viral antigens were also detected in considerable numbers of cuboidal, cytokeratin-positive cells, hence identified as type II pneumocytes (Fig. 1d and Supplementary Fig. 8), as has been reported for highly pathogenic avian H5N1 influenza viruses[6]. Upon infection with KUTK-4, large sections of infected lungs showed thickening of the alveolar wall on day 3 after infection (Fig. 1e). Although the infiltration of inflammatory cells was prominent at the alveolar wall (Fig. 1f), viral antigens were sparse and detected in type I (but not type II) pneumocytes (Fig. 1g). By contrast, the lungs of non-infected animals show clear alveolar spaces (Fig. 1h).

On day 7 after infection, lung pathology remained more severe for CA04- than for KUTK-4-infected lungs (Supplementary Fig. 7), although regenerative changes were seen for CA04. Nonetheless, considerable numbers of antigen-positive cells were still detectable (Supplementary Fig. 7c). Collectively, these findings demonstrate that CA04 causes more severe lung lesions in non-human primates than does a contemporary human influenza virus.

Induction of pro-inflammatory cytokines/chemokines in the lungs of CA04-infected macaques was variable at day 3 after infection (Supplementary Fig. 9). However, consistent with persisting lung pathology and inflammation on day 7 after infection, the levels of MCP-1, MIP-1α, IL-6 and IL-18 were markedly higher in the lungs of two of three CA04-infected macaques.

Ferrets are widely accepted as a suitable small-animal model for influenza virus pathogenicity and transmissibility studies. Infection of ferrets with S-OIVs or KUTK-4 did not cause marked changes in body temperature or weight in any group (data not shown). Although all test viruses were detected in nasal turbinates at similar titres on day 3 after infection (Supplementary Table 3), S-OIVs replicated to higher titres in trachea and lungs.

Pathological examination detected similar levels of viral antigen in the nasal mucosa of both CA04- and KUTK-4-infected ferrets (Supplementary Fig. 10a and e). However, the lungs of CA04-infected ferrets showed more severe bronchopneumonia with prominent viral antigen expression in the peribronchial glands and a few alveolar cells (Supplementary Fig. 10b–d) on day 3 after infection. By contrast, most of the lung appeared normal after infection with

Table 1 | Virus titres in organs of infected cynomolgus macaques

Organ	A/California/04/09 (H1N1)						A/Kawasaki/UTK-4/09 (H1N1)					
	Day 3 after infection			Day 7 after infection			Day 3 after infection			Day 7 after infection		
	1	2	3	4	5	6	7	8	9	10	11	12
Nasal mucosa	4.7	3.3	–	–	–	–	–	–	–	–	–	–
Oro/nasopharynx	6.3	4.4	4.7	–	7.9	–	–	–	4.3	–	–	4.8
Tonsil	6.4	–	–	–	7.1	–	–	–	2.8	–	–	3.0
Trachea	5.9	2.0	5.6	–	–	–	2.0	4.1	–	3.7	–	5.4
Bronchus (right)	5.7	2.9	4.3	–	5.1	–	–	2.5	–	3.5	–	3.8
Bronchus (left)	5.9	–	6.1	–	5.1	–	–	–	–	3.3	–	5.1
Lung (upper right)	5.7	5.6	4.5	–	–	–	2.7	–	–	–	–	–
Lung (middle right)	5.6	6.4	6.9	–	–	–	2.3	2.6	2.5	–	–	–
Lung (lower right)	6.1	4.5	6.0	–	–	–	2.6	2.6	–	–	–	3.4
Lung (upper left)	4.7	4.3	6.4	–	–	–	–	–	–	–	–	–
Lung (middle left)	5.8	4.3	6.3	–	–	–	–	–	–	–	–	–
Lung (lower left)	6.7	4.5	6.6	–	–	–	–	–	–	–	–	2.3
Conjunctiva	3.6	–	–	–	–	–	–	–	–	–	–	–

Cynomolgus macaques were inoculated with $10^{7.4}$ p.f.u. of virus (6.7 ml) through multiple routes (see Methods). Three macaques per group were killed on days 3 and 7 after infection for virus titration. No virus was recovered from lymph nodes (chest), heart, spleen, kidneys or liver of any of the animals. A dash indicates that virus was not detected (detection limit: $2 \log_{10}$ p.f.u. g^{-1}). Numbers 1 to 12 indicate animal identification number. Values indicate virus titre (mean $\log_{10}$ p.f.u. g^{-1}).
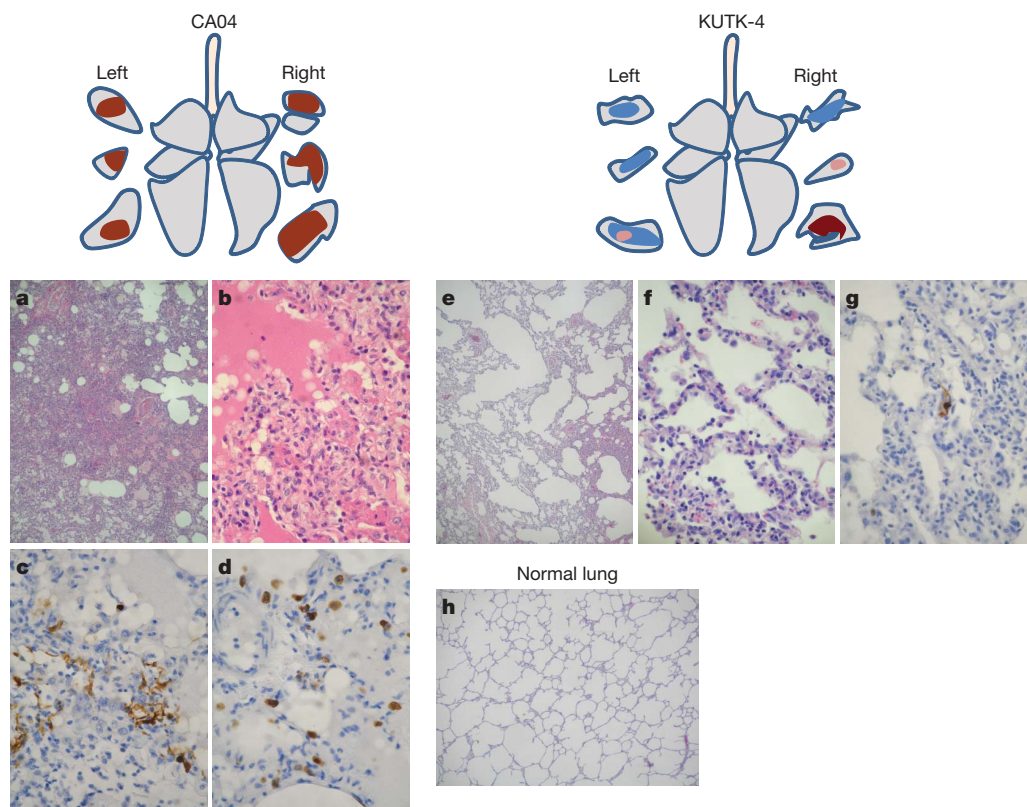
Figure 1 | Pathological examination of the lungs of infected cynomolgus macaques. a–h, Representative pathological images of CA04-infected (macaque no. 1, **a–d**), KUTK-4-infected (macaque no. 7, **e–g**) and mock-infected (**h**) lungs on day 3 after infection. One or two sections per lung lobe were examined. Representative findings are shown to depict the distribution of lesions in the sections (shown as cross-sections placed next to illustrations of each lung lobe), with or without viral antigen, as follows: brown, severe lung lesion containing moderate to many viral-antigen-positive cells; pink, mild lung lesions containing a few viral-antigen-positive cells; blue, lung lesions with alveolar wall thickening, with remaining air spaces unaffected. Original magnification: **a, e, h**, ×40; **b–d, f, g**, ×400.

KUTK-4 (Supplementary Fig. 10f and g). Thus, in all three mammalian models tested, CA04 seemed to be more pathogenic than a contemporary human H1N1 virus, KUTK-4.

Efficient human-to-human transmission is a critical feature of pandemic influenza viruses. To assess the transmissibility of CA04, naive ferrets in perforated cages were placed next to ferrets inoculated with 10^6 p.f.u. of CA04 (see Methods for detailed procedures). This experimental setting allows for aerosol transmission (that is, the exchange of respiratory droplets between the inoculated and non-inoculated ferrets) but prevents transmission by direct and indirect contact. All three contact ferrets were positive for CA04 virus on days 3 and 5 after infection (Supplementary Table 4). This transmission pattern is comparable to those of two human control influenza viruses that are known to transmit among ferrets: KUTK-4 and A/Victoria/3/75 (H3N2)[9]. By contrast, an avian influenza virus (A/duck/Alberta/35/76; H1N1) did not transmit (Supplementary Table 4).

Genetic analysis suggests that S-OIV originated in pigs[1]. However, there were no confirmed influenza virus outbreaks in Central American pigs before the reported S-OIV infections in humans. To assess S-OIV replication in pigs, we inoculated specific-pathogen-free miniature pigs, which are easier to manage, with CA04 or a classical swine influenza virus (A/swine/Hokkaido/2/81, H1N1). No signs of disease were observed (data not shown), although both viruses replicated efficiently in the respiratory organs of these animals (Supplementary Tables 5 and 6). Slightly higher titres of CA04 were detected in lungs on day 3 after infection, which is supported by pathological findings that show more apparent bronchitis and bronchiolitis in pigs infected with CA04 (Supplementary Fig. 11). The asymptomatic infection of CA04, despite efficient virus replication, might explain the lack of reports of S-OIV outbreaks in pigs before virus transmission to humans.

Antiviral compounds are the first line of defence against pandemic influenza viruses. Sequence analysis suggests that S-OIVs are resistant to ion channel inhibitors such as amantadine and rimantadine[1]. We therefore tested the licensed neuraminidase inhibitors oseltamivir and zanamivir, the experimental neuraminidase inhibitor R-125489 (the active form of CS-8958[10]) and the experimental compound T-705 (a broad-spectrum viral RNA polymerase inhibitor[11]) for their efficacy against CA04. In cell culture, CA04 was highly susceptible to all compounds tested (Supplementary Table 7), as were the human H1N1 control viruses A/Kawasaki/UTK-23/08 and KUTK-4, with the exception of the known oseltamivir resistance of KUTK-4. Comparable sensitivities were also found in an enzymatic neuraminidase inhibition assay[12] (Supplementary Table 8) and in mice (Fig. 2), consistent with observations in clinical settings.

A recent report suggested that 33% of individuals over 60 years of age had neutralizing antibodies to CA04 (http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5819a1.htm; Morbidity and Mortality Weekly Report, Centers for Disease Control and Prevention), probably due to previous exposure to antigenically similar H1N1 viruses. In fact, both the human H1N1 viruses that circulated until 1957 and the classical swine virus HA gene of S-OIVs are descendants of the 1918 pandemic virus, possibly explaining their antigenic relatedness. In 1977, H1N1 viruses re-emerged that were genetically and antigenically very closely related to viruses circulating in the 1950s[13] and should thus have elicited neutralizing antibodies to CA04 among younger age groups; however, this does not seem to be the case, according to the above described report. To resolve this puzzling finding, we assessed the neutralizing activities of sera collected from a broad range of age groups against CA04 and KUTK-4. We used two sets of donor sera, collected in 1999 from residents and workers in a nursing home (donor set 1), and in April
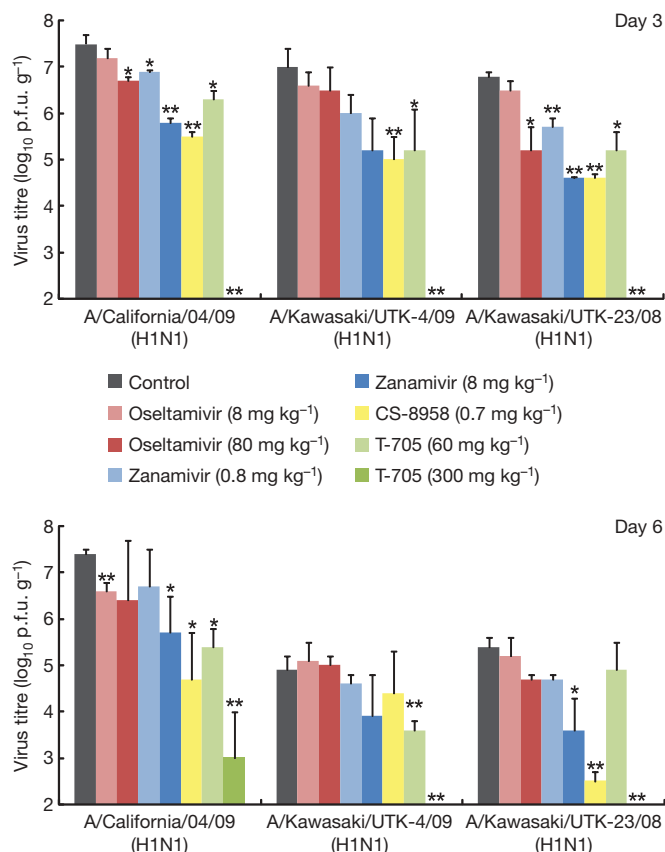
Figure 2 | CA04 sensitivity to antiviral compounds in mice. Mice were intranasally inoculated with 10^4 p.f.u. (50 μl) of CA04, KUTK-4 or A/Kawasaki/UTK-23/08 (H1N1). At 1 h after infection, mice were administered oseltamivir phosphate, zanamivir, CS-8958, T-705, or distilled water and PBS (control). Three mice per group were killed on days 3 and 6 after infection and the virus titres in lungs were determined by plaque assays in MDCK cells; results are reported as means ± s.d. The statistical significance of differences in lung virus titres of control mice and those treated with antivirals were assessed by use of the Student's t-test (asterisk, $P < 0.05$; double asterisk, $P < 0.01$).

2009 from workers and patients in a hospital (donor set 2). High neutralizing activity against KUTK-4 was detected for many sera in donor set 2 (Fig. 3), but not for sera in donor set 1, probably because these sera were collected before the emergence of the current human H1N1 viruses. Interestingly, with few exceptions, no appreciable neutralizing antibodies against CA04 were found for individuals born after 1920; however, many of those born before 1918 had high neutralizing antibody titres (individual neutralizing antibody titres are shown in Supplementary Table 9). These data indicate that infection with the 1918 pandemic virus or closely related human H1N1 viruses, but not infection with antigenically divergent human H1N1 viruses circulating in the 1920s to 1950s, and again since 1977, elicited neutralizing antibodies to S-OIVs.

Our findings indicate that S-OIVs are more pathogenic in mammalian models than seasonal H1N1 influenza viruses. In fact, the ability of CA04 to replicate in the lungs of mice, ferrets and non-human primates, and to cause appreciable pathology in this organ, is reminiscent of infections with highly pathogenic H5N1 influenza viruses[14], as acknowledged in a recent report by the World Health Organization (http://www.who.int/wer/2009/wer8421/en/index.html). We therefore speculate that the high replicative ability of S-OIVs might contribute to a viral pneumonia characterized by diffuse alveolar damage that contributes to hospitalizations and fatal cases where no other underlying health issues exist (http://www.who.int/wer/2009/wer8421/en/index.html). In addition, sustained person-to-person transmission might result in the emergence of more pathogenic variants, as observed with
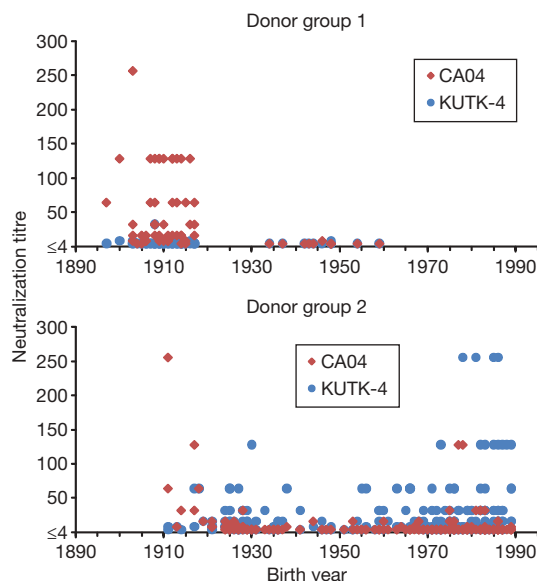
Figure 3 | Neutralization activities in human sera against viruses. Human sera of donor groups 1 (collected in 1999) and 2 (collected in April and May of 2009) were subjected to neutralization assays with CA04 and KUTK-4. Because the sera of donor group 1 were collected in 1999, little neutralization activity was expected against KUTK-4, which was isolated in 2009.

the 1918 pandemic virus (reviewed in ref. 15). Furthermore, S-OIVs may acquire resistance to oseltamivir through mutations in their NA gene (as recently witnessed with human H1N1 viruses[16]), or through reassortment with co-circulating, oseltamivir-resistant seasonal human H1N1 viruses. Collectively, our findings are a reminder that S-OIVs have not yet garnered a place in history, but may still do so, as the pandemic caused by these viruses has the potential to produce a significant impact on human health and the global economy.

METHODS SUMMARY

Viruses and cells. All swine-origin H1N1 viruses were isolated and passaged in MDCK cells to produce viral stocks. The viruses and their passage histories are described in Methods. All experiments with S-OIVs were performed in approved enhanced biosafety level 3 (BSL3) containment laboratories.

MDCK cells and MDCK cells overexpressing the β-galactoside α2,6-sialyltransferase I gene[17] were maintained in Eagle's minimal essential medium (MEM) containing 5% newborn calf serum. Human airway epithelial (HAE) cells were obtained from residual surgical tissue trimmed from lungs during the process of transplantation. The bronchial specimens were dissected and enzymatically digested, and monolayers of HAE cells were isolated, cultured and differentiated as previously described[18].

Animals. Five- and six-week-old female BALB/c mice (Jackson Laboratory and Japan SLC Inc.), approximately three-to-four-year-old cynomolgus macaques (Ina Research Inc.), five-to-eight-month-old male ferrets (Marshall Farms and Triple F Farms) and two-month-old female specific-pathogen-free miniature pigs (Nippon Institute for Biological Science) were used according to approved protocols for the care and use of animals. Detailed procedures are provided in Methods.

Antiviral sensitivity of viruses in mice. Five-week-old female BALB/c mice (Japan SLC Inc.; groups of six) were anaesthetized with sevoflurane and intranasally inoculated with 10^4 p.f.u. (volume, 50 μl) of CA04, KUTK-4, or A/Kawasaki/UTK-23/08 (H1N1). At 1 h after infection, mice were administered antiviral compounds as described in detail in Methods. Three mice per group were killed on days 3 or 6 after infection and the virus titres in lungs were determined by plaque assays in MDCK cells.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N. Engl. J. Med.* **360**, 2605–2615 (2009).

2.   Kawaoka, Y. & Webster, R. G. Sequence requirements for cleavage activation of influenza virus hemagglutinin expressed in mammalian cells. *Proc. Natl Acad. Sci. USA* **85**, 324–328 (1988).

3.   Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* **293**, 1840–1842 (2001).

4.   Iwakura, Y., Nakae, S., Saijo, S. & Ishigame, H. The roles of IL-17A in inflammatory immune responses and host defense against pathogens. *Immunol. Rev.* **226**, 57–79 (2008).

5.   Hamada, H. *et al.* Tc17, a unique subset of CD8 T cells that can protect against lethal influenza challenge. *J. Immunol.* **182**, 3469–3481 (2009).

6.   Baskin, C. R. *et al.* Early and sustained innate immune response defines pathology and death in nonhuman primates infected by highly pathogenic influenza virus. *Proc. Natl Acad. Sci. USA* **106**, 3455–3460 (2009).

7.   Rimmelzwaan, G. F. *et al.* Pathogenesis of influenza A (H5N1) virus infection in a primate model. *J. Virol.* **75**, 6687–6691 (2001).

8.   Kobasa, D. *et al.* Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. *Nature* **445**, 319–323 (2007).

9.   Maines, T. R. *et al.* Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model. *Proc. Natl Acad. Sci. USA* **103**, 12121–12126 (2006).

10. Yamashita, M. *et al.* CS-8958, a prodrug of the new neuraminidase inhibitor R-125489, shows long-acting anti-influenza virus activity. *Antimicrob. Agents Chemother.* **53**, 186–192 (2009).

11. Furuta, Y. *et al. In vitro* and *in vivo* activities of anti-influenza virus compound T-705. *Antimicrob. Agents Chemother.* **46**, 977–981 (2002).

12. Hayden, F. G. *et al.* Inhaled zanamivir for the prevention of influenza in families. Zanamivir Family Study Group. *N. Engl. J. Med.* **343**, 1282–1289 (2000).

13. Nakajima, K., Desselberger, U. & Palese, P. Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature* **274**, 334–339 (1978).

14. Peiris, J. S. *et al.* Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet* **363**, 617–619 (2004).

15. Wright, P. F., Neumann, G. & Kawaoka, Y. *Fields Virology* (eds Knipe, D. M. *et al.*) 1691–1740 (Wolters Kluwer/Lippincott Williams & Wilkins, 2007).

16. Moscona, A. Global transmission of oseltamivir-resistant influenza. *N. Engl. J. Med.* **360**, 953–956 (2009).

17. Hatakeyama, S. *et al.* Enhanced expression of an α2,6-linked sialic acid on MDCK cells improves isolation of human influenza viruses and evaluation of their sensitivity to a neuraminidase inhibitor. *J. Clin. Microbiol.* **43**, 4139–4146 (2005).

18. Jakiela, B., Brockman-Schneider, R., Amineva, S., Lee, W. M. & Gern, J. E. Basal cells of differentiated bronchial epithelium are more susceptible to rhinovirus infection. *Am. J. Respir. Cell Mol. Biol.* **38**, 517–523 (2008).

Author Contributions Y.I., K.S., M.K., T.W., Y.S., M.H., Y.M., D.T., Y.S.-T., T.N., M. Imai, S.W., K.I.-H., T.H., N.S., H.K., K.O. and Y.K. designed the experiments; Y.I., K.S., M.K., T.W., Y.S., M.H., D.T., Y.S.-T., T.N., S.S., M. Imai, Y.H., S.W., C.L., S.Y., K.F., S.M., H. Imai, S.K., M. Ito, R.T., K.I.-H., M.S., T.H., Kei Takahashi, A.M., H. Ishigaki, M. Nakayama, M. Okamatsu, Kazuo Takahashi, D.W., P.A.S., R.S., H.S., Y.F., M. Yamashita, K.M., K.N., M. Nakamura, R.B.-S., J.G., H.M. and M. Yamazaki performed the experiments; Y.I., K.S., M.K., T.W., Y.S., M.H., Y.M., Y.S.-T., T.N., M. Imai, S.W., C.L., S.Y., K.I.-H., T.H., H.G., M.S., M. Ozawa, G.N., H.K., K.O. and Y.K. analysed data; Y.I., K.S., M.K., T.W., Y.S., M.H., Y.M., Y.S.-T., T.N., M. Imai, K.I.-H., M.S., M. Ozawa, G.N., K.O. and Y.K. wrote the manuscript. Y.I., K.S., M.K., T.W., Y.S., M.H. and Y.M. contributed equally to this work.

METHODS

Viruses. A/California/04/09 (H1N1; CA04) was provided by the Centers for Disease Control (CDC). A/Wisconsin/WSLH049/09 (H1N1) was isolated from a patient with mild symptoms, whereas A/Wisconsin/WSLH34939/09 (H1N1) was isolated from a hospitalized patient. A/Netherlands/603/09 (H1N1) was isolated from a patient with mild symptoms and was provided by R. Fouchier. A/Osaka/164/09 (H1N1) was also isolated from a patient with mild symptoms.

The following influenza viruses served as controls: A/Kawasaki/UTK-4/09 (H1N1; KUTK-4; passaged twice in MDCK cells), an oseltamivir-resistant seasonal human virus; A/WSN/33 (H1N1; generated by reverse genetics and passaged twice in MDCK cells), a typical spherical influenza virus[19]; A/Kawasaki/UTK-23/08 (H1N1; passaged twice in MDCK cells), an oseltamivir-sensitive seasonal human virus; A/Victoria/3/75 (H3N2; passaged several times in eggs after it was obtained from the CDC), a human virus; A/swine/Hokkaido/2/81 (H1N1; passaged several times in eggs), a classical swine virus; and A/duck/Alberta/35/76 (H1N1; passaged several times in eggs), an avian virus. All experiments with S-OIV viruses were performed in enhanced biosafety level 3 (BSL3) containment laboratories at the University of Wisconsin-Madison, which are approved for such use by the CDC and the US Department of Agriculture, or in BSL3 containment laboratories at the University of Tokyo, the Shiga University of Medical Science, or the Hokkaido University, all of which are approved for such use by the Ministry of Agriculture, Forestry and Fisheries, Japan.

Viral pathogenesis in mice. Six-week-old female BALB/c mice (Jackson Laboratory and Japan SLS Inc.) were used in this study. Baseline body weights were measured before infection. Three mice per group were anaesthetized with isoflurane and intranasally inoculated with 10^2, 10^3, 10^4, or 10^5 p.f.u. (50 μl) of CA04 and KUTK-4, or undiluted virus from virus stocks (CA04, $10^{6.5}$ p.f.u.; KUTK-4, $10^{6.6}$ p.f.u.). Body weight and survival were monitored daily for 14 days and mice with body weight loss of more than 25% of pre-infection values were killed. For virological and pathological examinations, 6 mice per group were intranasally infected with 10^5 p.f.u. of S-OIVs and KUTK-4 and 3 mice per group were killed on days 3 and 6 after infection. The virus titres in various organs were determined by plaque assays in MDCK cells.

Growth kinetics of virus in human airway epithelial (HAE) cells. Cultures of differentiated HAE cells were washed extensively with PBS to remove accumulated mucus and infected with virus at a multiplicity of infection (MOI) of 0.001 from the apical surface. The inoculum was removed after 1 h of incubation at 35 °C, and cells were further incubated at 35 °C. Samples were collected at 12, 24, 48, 72 and 96 h after infection from the apical surface. Apical harvesting was performed by adding 500 μl of medium to the apical surface, followed by incubation for 30 min at 35 °C, and removal of the medium from the apical surface. The titres of viruses released into the cell culture supernatant were determined by plaque assay in MDCK cells.

Experimental infection of cynomolgus macaques. Approximately three-to-four-year-old cynomolgus macaques (*Macaca fascicularis*) from the Philippines (obtained from Ina Research Inc.), weighing 2.1–3.0 kg and serologically negative by AniGen AIV antibody ELISA, which detects all influenza A virus subtypes (Animal Genetics Inc.), were used in this study. Baseline body weights were established by two or three measurements before infection. Under anaesthesia, telemetry probes (TA10CTA-D70, Data Sciences International) were implanted in the peritoneal cavities of animals to monitor body temperature. Six macaques per group were intramuscularly anaesthetized with ketamine (5 mg per kg) and xylazine (1 mg per kg) and inoculated with a suspension containing $10^{6.5}$ p.f.u. ml^{-1} of CA04 or KUTK-4 virus through a combination of intratracheal (4.5 ml), intranasal (0.5 ml per nostril), ocular (0.1 ml per eye) and oral (1 ml) routes (resulting in a total infectious dose of $10^{7.4}$ p.f.u.). Macaques were monitored every 15 min for changes in body temperature. On days 1, 3, 5 and 7 after infection, nasal and tracheal swabs and bronchial brush samples were collected. On days 3 and 7 after infection, 3 macaques per group were killed for virological and pathological examinations. The virus titres in various organs and swabs were determined by plaque assays in MDCK cells. Experiments were carried out in accordance with the Guidelines for the Husbandry and Management of Laboratory Animals of the Research Center for Animal Life Science at Shiga University of Medical Science, Shiga, Japan, and approved by the Shiga University of Medical Science Animal Experiment Committee and Biosafety Committee.

Experimental infection of ferrets. We used five-to-eight-month-old male ferrets (Marshall Farms and Triple F Farms), which were serologically negative by haemagglutination inhibition (HI) assay for currently circulating human influenza viruses. Baseline body temperatures and body weights were established by one or two measurements before infection. Six ferrets per group were intramuscularly anaesthetized with ketamine and xylazine (5 mg and 0.5 mg per kg of body weight, respectively) and intranasally inoculated with 10^6 p.f.u. (500 μl) of

S-OIVs or KUTK-4. On days 3 and 6 after infection, 3 ferrets per group were killed for virological and pathological examinations. The virus titres in various organs were determined by plaque assays in MDCK cells.

Experimental infection of miniature pigs. Two-month-old female specific-pathogen-free miniature pigs (Nippon Institute for Biological Science), which were serologically negative by AniGen AIV antibody ELISA for currently circulating influenza viruses, were used in this study. Baseline body temperatures were measured once before infection. Four pigs per group were intranasally inoculated with $10^{6.2}$ p.f.u. (1 ml) of viruses. Nasal swabs were collected daily. On day 3 after infection, two pigs per group were killed and their tissues collected for examination. On day 14 after infection, the remaining two pigs per group were killed for virological and pathological examinations. Virus titres in various organs and swabs were determined by plaque assays in MDCK cells. The miniature pigs used in this study were housed in self-contained isolator units (Tokiwa Kagaku) at a BSL3 facility and experiments were conducted in accordance with guidelines established by the Animal Experiment Committee of the Graduate School of Veterinary Medicine, Hokkaido University, Japan.

Pathological examination. Excised tissues of the nasal turbinates, trachea and/or lungs of killed mice, macaques, ferrets and pigs were preserved in 10% phosphate-buffered formalin. Tissues were then processed for paraffin embedding and cut into 5-μm-thick sections. One section from each tissue sample was stained using a standard haematoxylin-and-eosin procedure, whereas another one was processed for immunohistological staining with an anti-influenza virus rabbit antibody (R309; prepared in our laboratory) that reacts comparably with CA04 and KUTK-4. Specific antigen–antibody reactions were visualized by 3,3′-diaminobenzidine tetrahydrochloride staining using a Dako EnVision system (Dako Co. Ltd).

Ferret transmission study. For transmission studies in ferrets, animals were housed in adjacent transmission cages that prevent direct and indirect contact between animals but allow spread of influenza virus through the air. Three or two 5-to-8-month-old ferrets were intranasally inoculated with 10^6 p.f.u. (500 μl) of CA04, KUTK-4, A/Victoria/3/75 (H3N2) or A/duck/Alberta/35/76 (H1N1) (inoculated ferrets). One day after infection, three or two naive ferrets were each placed in a cage adjacent to an inoculated ferret (contact ferrets). All ferrets were monitored daily for changes in body temperature and weight, and the presence of clinical signs. To assess viral replication in the upper respiratory tract, viral titres were determined in nasal washes collected from virus-inoculated and contact ferrets on day 1 after inoculation or co-housing, respectively, and then every other day (up to 9 days).

Cytokine and chemokine measurement. For cytokine and chemokine measurement, homogenates of mouse lungs were processed with the Bio-Plex Mouse Cytokine 23-Plex and 9-Plex panels (Bio-Rad Laboratories), whereas macaque lung homogenates were measured with the MILLIPLEX MAP Non-human Primate Cytokine/Chemokine Panel–Premixed 23-Plex (Millipore). Array analysis was performed by Bio-Plex Protein Array system (Bio-Rad Laboratories).

Antiviral sensitivity of viruses in mice. To test the antiviral sensitivity of viruses in mice, animals were infected as described in the Methods Summary section and 1 h later administered the following antiviral compounds: (1) oseltamivir phosphate: 8 or 80 mg per kg per 400 μl (divided into two oral administrations per day) for 5 days; (2) zanamivir: 0.8 or 8 mg per kg per 50 μl in one daily intranasal administration for 5 days; (3) CS-8958: 0.7 mg per kg per 50 μl in one intranasal administration; (4) T-705: 60 or 300 mg per kg per 400 μl (divided into two oral administrations per day) for 5 days; (5) or distilled water orally (200 μl) and PBS intranasally (50 μl). Three mice per group were killed on day 3 or 6 after infection and the virus titres in lungs were determined by plaque assays in MDCK cells.

Sensitivity to antiviral compounds in tissue culture. MDCK cells overexpressing the β-galactoside α2,6-sialyltransferase I gene (or, for studies with T-705, regular MDCK cells) were infected with CA04, KUTK-4, or A/Kawasaki/UTK-23/08 (H1N1) at a multiplicity of infection of 0.001. After incubation for 1 h at 37 °C, growth medium containing various concentrations of oseltamivir carboxylate (the active form of oseltamivir), zanamivir, R-125489 (the active form of CS-8958), or T-705 was added to the cells. Twenty-four hours later, the culture supernatants were harvested and the 50% tissue-culture infectious dose (TCID$_{50}$) in MDCK cells determined. On the basis of the TCID$_{50}$ value, the 90% inhibitory concentration (IC$_{90}$) was calculated.

Neuraminidase inhibition assay. To assess the sensitivity of viruses to neuraminidase inhibitors (that is, oseltamivir, zanamivir and CS-8958), neuraminidase inhibition assays were performed as described previously[20]. Briefly, diluted viruses were mixed with various concentrations of oseltamivir carboxylate, zanamivir, or R-125489 in 2-(N-morpholino)ethanesulphonic acid containing calcium chloride, and incubated for 30 min at 37 °C. Then, we added methylumbelliferyl-N-acetylneuraminic acid (Sigma) as a fluorescent substrate to this mixture. After incubation for 1 h at 37 °C, sodium hydroxide in 80% ethanol was added to the mixture to stop the reaction. The fluorescence of the solution was measured at an

excitation wavelength of 360 nm and an emission wavelength of 465 nm and the 50% inhibitory concentration (IC_{50}) was calculated.

Neutralization assay with human sera. Human sera were collected in 1999 or 2009 from donor group 1 (age range: 50–112 years as of 2009, mean = 92.7 ± 15.0 years) or 2 (age range: 20–68 years as of 2009, mean = 48.2 ± 23.7 years), respectively. These sera were treated with receptor-destroying enzyme (DENKA SEIKEN CO.) to remove inhibitors of influenza virus replication. One hundred $TCID_{50}$ (50% tissue culture infectious dose) of CA04 and KUTK-4 were pre-incubated with twofold serial dilutions of treated sera, incubated for 60 min on MDCK cells, which were then observed for cytopathic effects to determine the neutralizing activity of the test sera. Our research protocol was approved by the Research Ethics Review Committee of the Institute of Medical Science, the University of Tokyo (approval numbers: 21-6-0428 for donor group 1; 21-7-0529 for donor group 2).

Immunofluorescence microscopy. MDCK cells were infected with CA04, KUTK-4, or WSN and fixed with 4% paraformaldehyde 16–24 h later.

Infected cells were incubated with the following primary antibodies: mouse anti-HA (7B1b), anti-HA (IVC102), or mouse anti-HA (WS3-54) antibody against CA04, KUTK-4 or WSN, respectively. Cells were then incubated with Alexa Fluor 488 goat anti-mouse immunoglobulin G (Invitrogen), and examined with a confocal laser-scanning microscope (LSM510META; Carl Zeiss).

Electron microscopy. MDCK cells were infected with CA04, KUTK-4 or WSN at a multiplicity of infection of 10. At 16–24 h after infection, cells were processed for ultrathin section electron microscopy and scanning electron microscopy as described previously[19,21].

19. Noda, T. *et al.* Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* **439**, 490–492 (2006).
20. Kiso, M. *et al.* Resistant influenza A viruses in children treated with oseltamivir: descriptive study. *Lancet* **364**, 759–765 (2004).
21. Neumann, G. *et al.* Ebola virus VP40 late domains are not essential for viral replication in cell culture. *J. Virol.* **79**, 10300–10307 (2005).

LETTERS

The ethylene response factors *SNORKEL1* and *SNORKEL2* allow rice to adapt to deep water

Yoko Hattori[1], Keisuke Nagai[1], Shizuka Furukawa[1], Xian-Jun Song[1], Ritsuko Kawano[3], Hitoshi Sakakibara[4], Jianzhong Wu[5], Takashi Matsumoto[5], Atsushi Yoshimura[3], Hidemi Kitano[1], Makoto Matsuoka[1], Hitoshi Mori[2] & Motoyuki Ashikari[1]

Living organisms must acquire new biological functions to adapt to changing and hostile environments. Deepwater rice has evolved and adapted to flooding by acquiring the ability to significantly elongate its internodes, which have hollow structures and function as snorkels to allow gas exchange with the atmosphere, and thus prevent drowning[1–3]. Many physiological studies have shown that the phytohormones ethylene, gibberellin and abscisic acid are involved in this response[4–8], but the gene(s) responsible for this trait has not been identified. Here we show the molecular mechanism of deepwater response through the identification of the genes *SNORKEL1* and *SNORKEL2*, which trigger deepwater response by encoding ethylene response factors involved in ethylene signalling. Under deepwater conditions, ethylene accumulates in the plant and induces expression of these two genes. The products of *SNORKEL1* and *SNORKEL2* then trigger remarkable internode elongation via gibberellin. We also demonstrate that the introduction of three quantitative trait loci from deepwater rice into non-deepwater rice enabled the latter to become deepwater rice. This discovery will contribute to rice breeding in lowland areas that are frequently flooded during the rainy season.

Deepwater rice has a unique ability to elongate its internodes with increasing water depth (Fig. 1a, Supplementary Movies 1, 2, and Supplementary Fig. 1). We attempted to identify genes that regulate deepwater responses using quantitative trait loci (QTL) analysis combined with positional cloning. The non-deepwater rice variety Taichung65 (T65) shows little elongation under deepwater conditions (Fig. 1b left panel, 1c). In contrast, the deepwater rice C9285 shows significant elongation of leaves and internodes under deep water (Fig. 1b middle, 1c). Using the progeny of T65/C9285, we previously detected three major QTLs for deepwater response on chromosomes 1, 3 and 12 (refs 9, 10; Fig. 1b middle), and also detected these QTLs in the same chromosome regions in different populations[11,12]. Among these, the QTL on chromosome 12 was the most effective for deepwater response[9–12]. To evaluate the effects of the QTL on chromosome 12, we developed the nearly isogenic line 12 (NIL-12), which possesses the C9285 genomic fragment on chromosome 12 in a T65 genetic background (Supplementary Fig. 2a). NIL-12 responded to deep water by internode elongation (Fig. 1b right, 1c).

Positional cloning and gain-of-function analysis using a transgenic approach allowed us to identify two genes, *SNORKEL1* (*SK1*) and *SNORKEL2* (*SK2*), that regulated the deepwater response (Supplementary Information, Supplementary Figs 3–5, Supplementary Table 1). The deepwater rice C9285 possesses *SK1* and *SK2*, although both genes are absent from the non-deepwater rice T65 (Supplementary Fig. 3c). *SK1* and *SK2* possess a putative nuclear localization signal and a single

APETALA2/ethylene response factor (AP2/ERF) domain (Fig. 1d, Supplementary Fig. 6). On the basis of phylogenetic analysis using the AP2/ERF domain, the *SK* genes were categorized as belonging to the ERF subfamily (Supplementary Fig. 7). The ERF domains in the *SK* genes showed a high similarity to those of *Arabidopsis thaliana (At)ERF1*, *Oryza sativa (Os)ERF1* and *SUB1A-1* (Fig. 1e, Supplementary Table 2)[13–15]. However, other than the AP2/ERF domain, neither *SK1* nor *SK2* showed similarity to any other known genes. The *SK* genes were significantly expressed under deepwater conditions, whereas these expressions were low under dry conditions in C9285 (Fig. 1f). Furthermore, the *SK* genes were expressed in leaf blade, leaf sheath, and basal parts of the stem, including nodes and internodes, in which the deepwater response occurs (Fig. 1g). We also generated transgenic plants that overproduced *SK* genes driven by the *OsAct1* promoter in T65 background (Supplementary Table 1). Compared to vector control plants, *SK1*-overproducers elongated one to three internodes and *SK2*-overproducers elongated one to seven internodes, even under dry conditions (Fig. 1h). These results suggest that the *SK* genes regulate internode elongation.

Subcellular localization analysis using green fluorescent protein (GFP) showed that the SK proteins localized to nuclei (Fig. 2a). In addition, analysis using the yeast one-hybrid system showed that SK proteins have transcription-activating abilities (Fig. 2b, Supplementary Information). These results suggest that *SK* genes have the potential to act as transcription factors. As *SK* genes encode ERF, we investigated their involvement in ethylene signalling. Ethylene and five other basic phytohormones—that is, gibberellin (GA), brassinosteroid (BR), auxin (IAA), cytokinin (CK) and abscisic acid (ABA)—were applied to C9285. Of these, only ethylene significantly upregulated the expression of the *SK* genes (Fig. 2c). Thus, *SK1* and *SK2* are responsive to ethylene.

Previous studies in *Arabidopsis* have indicated that the *EIN3* (*ETHYLENE-INSENSITIVE 3*) gene product activates *AtERF1* expression by binding to the *AtERF1* promoter in the ethylene signalling pathway[16]. We analysed the binding of the rice *EIN3*-like gene product (OsEIL1b, Supplementary Fig. 8) to the promoter regions of *SK1* and *SK2* (Supplementary Information). Electrophoretic mobility shift assays (EMSAs) showed that recombinant OsEIL1b protein bound to the *SK* promoter regions (Fig. 2d, Supplementary Table 3a, Supplementary Fig. 9, Supplementary Information). These two regions contained the target core sequences of the EIN3-binding site (Supplementary Table 3b)[16]. These results indicate that *SK1* and *SK2* are novel *ERFs* involved in the deepwater response of deepwater rice.

Kende *et al.* reported that application of ethylene induces internode elongation in deepwater rice[3]. We also investigated the relationship between ethylene and deepwater response. Application of

[1]Bioscience and Biotechnology Center, [2]Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan. [3]School of Agriculture, Kyushu University, Fukuoka 812-8581, Japan. [4]Plant Science Center, RIKEN, Yokohama 230-0045, Japan. [5]National Institute of Agrobiological Resources, Tsukuba 305-8602, Japan.
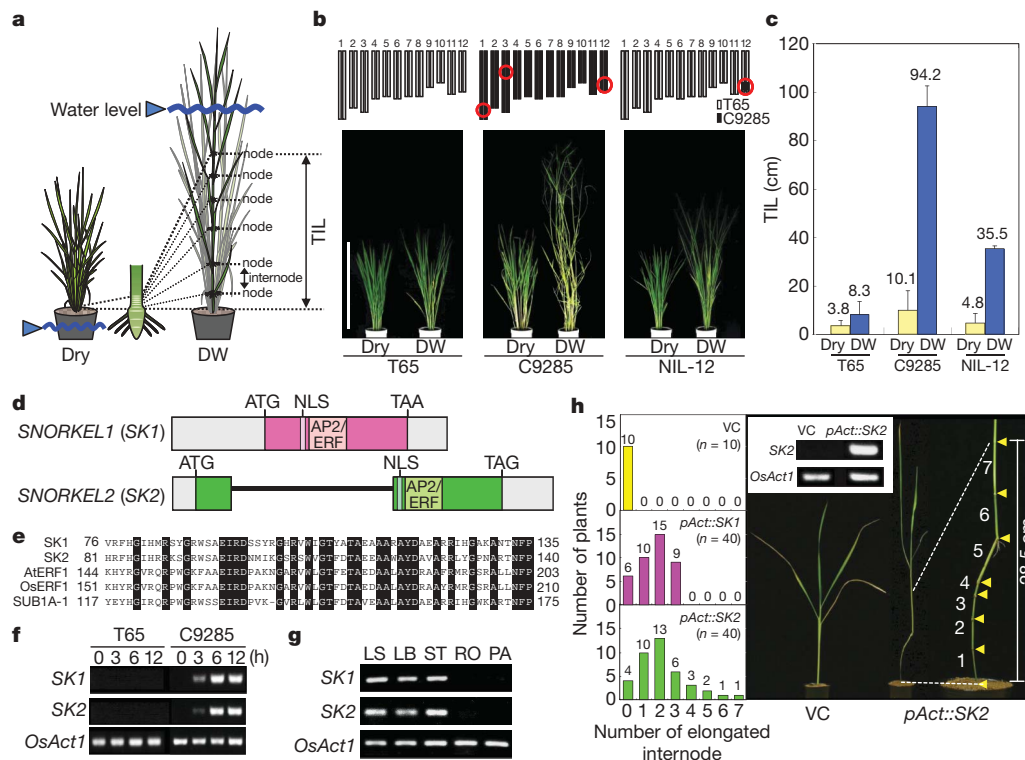
Figure 1 | Identification of genes responsible for deepwater response in rice. a, Image of deepwater response. Dry, water level is under the soil surface; DW, deepwater condition; TIL, total internode elongation length. **b**, Deepwater response in T65, C9285 and NIL-12. Red circles, positions of major QTLs[9,10]. Scale bar, 1 m. **c**, Quantitative evaluation of deepwater response in T65, C9285 and NIL-12. Data are mean + s.d., $n = 5$. **d**, Genome structures of SK1 and SK2. Pink and green boxes, exons. White boxes, 3′ or 5′ non-coding regions. NLS, nuclear localization signal. **e**, AP2/ERF domain in SKs. Black shading, identical amino acid residues. **f**, Expression analysis of SK genes in T65 and C9285. **g**, Organ-specific expression of SK genes. LS, leaf sheath; LB, leaf blade; ST, basal part of stem; RO, root; PA, panicle. **h**, Internode elongation in transgenic plants overproducing SK genes in T65 background under dry conditions. VC, vector control; pAct::SK1, overproducing SK1; pAct::SK2, overproducing SK2. Ten-leaf-stage plants were submerged for two weeks (**b**, **c**), for indicated times (**f**), or for 12 h (**g**). The rice actin gene (OsAct1) was used as a control (**f**, **g**).

ethylene enhanced internode elongation of C9285, whereas no such effect was observed in T65 (Fig. 2e). In addition, application of ethylene along with its inhibitor 1-methylcyclopropene (1-MCP) suppressed internode elongation. We also compared the internode elongation of C9285 in deepwater conditions with or without 1-MCP pretreatment (Fig. 2f). The internode elongation of plants pretreated with 1-MCP under deepwater conditions was less than in those without pretreatment. These results indicate that ethylene is involved in deepwater response. Ethylene is known to accumulate in deepwater rice under deepwater conditions[4]. We examined whether ethylene accumulation is a common phenomenon for both deepwater and non-deepwater rice (Supplementary Fig. 10). Under deepwater conditions, ethylene content increased about 25-fold within 24 h in both T65 and C9285 (Fig. 2g). This observation indicates that ethylene accumulation under deepwater conditions is a common response of both deepwater and non-deepwater rice. The ethylene diffusion rate in water is 1/10,000 of that in air[17]. Both ethylene production and its low diffusion rate in deep water would contribute to ethylene accumulation in rice.

The above results suggest a molecular mechanism for the deepwater rice response regulated by SK1 and SK2. Under deepwater conditions, ethylene accumulates in both deepwater and non-deepwater rice. The accumulated ethylene triggers the expression of the SK genes, leading to the induction of internode elongation in deepwater rice. In contrast, as non-deepwater rice T65 lacks SK genes, it shows no response to deepwater conditions.

Next, we examined the phytohormone contents of T65 and C9285 before and after submergence using liquid chromatography/mass spectrometry (LC/MS). Among the plant hormones investigated, GA_1 is the active GA, and its level in nodes was elevated after submergence of C9285, whereas the GA_1 content in T65 did not show any

marked increase before or after submergence (Fig. 3a). The concentrations of other phytohormones were not significantly different between T65 and C9285 (Supplementary Table 4, Supplementary Information). GA is known as the plant growth hormone[18]. Our results suggest that GA may be involved in the deepwater response. To examine this, we investigated the GA response in deepwater rice (Fig. 3b). In the presence of the GA biosynthetic inhibitor uniconazole, internode elongation under deepwater conditions was repressed in C9285, while internode elongation was induced without uniconazole. Application of GA induced internode elongation in C9285 under dry conditions. In addition, the repression of internode elongation by uniconazole was rescued by application of GA. These results supported the findings of previous physiological experiments that GA is involved in the deepwater response[5]. Signalling by SK genes may be directly or indirectly connected to GA, which then induces internode elongation.

We also investigated the evolution of SK genes. The wild rice species *Oryza rufipogon* (W0120) and *Oryza nivara* (W0106) are progenitors of *O. sativa*[19]. We examined the deepwater responses of W0120 and W0106, and found that W0120 responded dramatically to deepwater conditions with elongated internodes, whereas W0106 showed little response (Fig. 4a). We compared the sequences of the SK genes in these lines, using clones selected from bacterial artificial chromosome (BAC) libraries. W0120 possessed both SK1 and SK2, whereas W0106 possessed SK1 and had a new stop-codon in exon 2 of SK2 by insertion of a transposon (Fig. 4b). We detected the same stop-codon in other accessions of *O. nivara* (IRGC105319, IRGC105703, Supplementary Information). Gain-of-function analysis of the SK genes suggested that SK2 has a more pronounced effect than SK1 (Supplementary Information, Supplementary Fig. 5). Furthermore, recent research suggests that some *O. rufipogon* are adapted to wet areas and some *O. nivara* are adapted to dry areas[20]. This mutation in SK2 would not
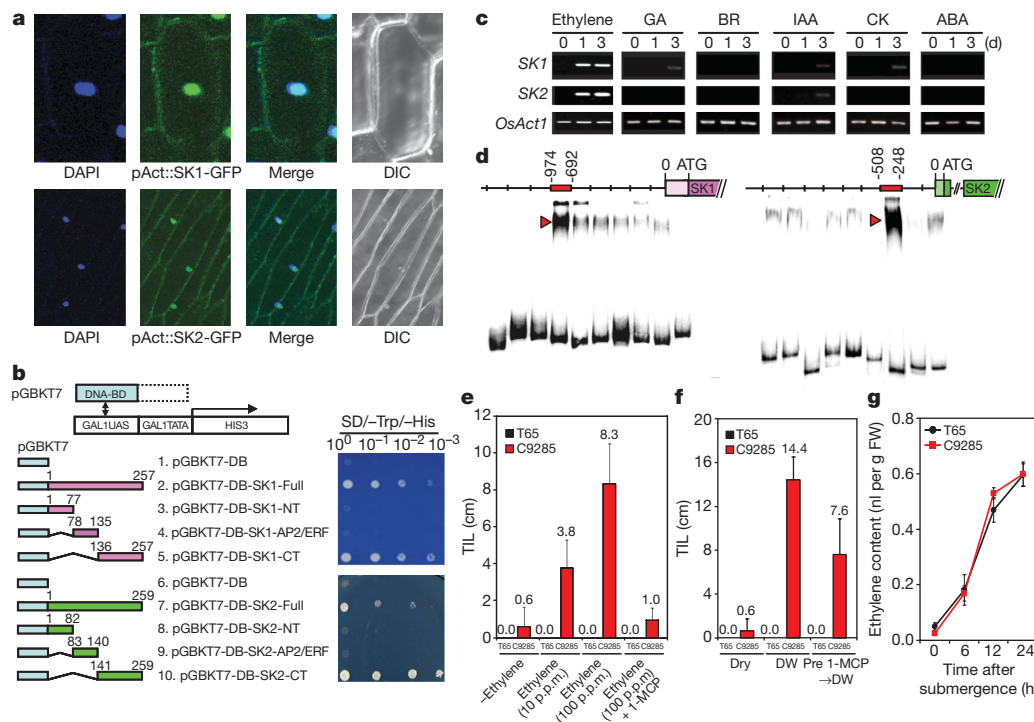
Figure 2 | Molecular characterization of *SK1* and *SK2*. a, Subcellular localization of SK1 and SK2 in onion epidermal cells. pAct::SK1-GFP and pAct::SK2-GFP activity was visible in nuclei. **b,** Transactivation activity assay in yeast. The corresponding full-length cDNA (full), N-terminal (NT), C-terminal (CT), or AP2/ERF domains alone (AP2/ERF) of *SK1* and *SK2* were fused to the GAL4-DNA binding domain (BD). The amino acid numbers are indicated above the fragment bars. **c,** Relationships between plant hormones and induction of *SK* genes. 10 p.p.m. ethylene, 10 μM GA, 10 nM BR, 20 μM IAA, 1 μM CK, 100 μM ABA were used. **d,** Sequence-specific binding of the recombinant OsEIL1b protein to *SK* promoters. Arrowheads indicate shifted bands. **e,** Ethylene promotes internode elongation without deepwater conditions in deepwater rice. Data are mean + s.d., $n = 8$. **f,** Internode elongation under deepwater conditions with ethylene inhibitor, 1-MCP (10 p.p.m.). Pre 1-MCP→DW, plants were pretreated by 1-MCP for 3 days before deepwater treatment; mean + s.d., $n = 8$. **g,** Measurement of ethylene content under deepwater conditions. g FW, grams fresh weight. Mean ± s.d., $n = 8$. Six-leaf-stage plants were subjected to each treatment for one week (**e**, **f**) or for the indicated time (**g**).

distinguish between *O. rufipogon* and *O. nivara*. However, in dry areas, *SK2* may be dispensable, and adaptation to such environment may have resulted in the loss of *SK2* in W0106. The wild rice species *Oryza glumaepatula* (IRGC105668), which grows in areas along the Amazon River in South America that are flooded during the rainy season, responds to deepwater conditions (Fig. 4a), and sequence analysis of its BAC clones showed that it possesses *SK2* and *SK2*-like genes, but not *SK1* (Fig. 4b). These results suggest that *SK2* is important for the deepwater response. As some wild rice species possess the *SK* genes, these genes may have been acquired before or during wild rice species divergence.
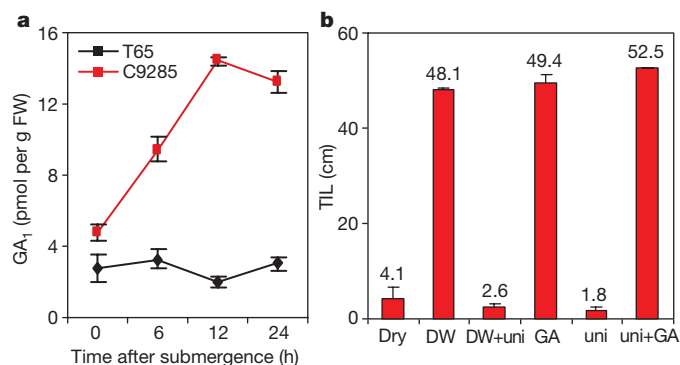
Figure 3 | GA response and molecular mechanism of deepwater response. a, GA$_1$ content in C9285 under deepwater conditions. Mean ± s.d., $n = 4$. **b,** GA responsiveness in C9285. Ten-leaf-stage plants were treated with 100 μM GA$_3$ with or without 1 μM uniconazole (uni) for one week. Mean ± s.d., $n = 8$.

Deepwater rice is mainly cultivated in lowland areas, and it is the only crop that can be grown in flooded areas in the rainy season. One problem with the cultivation of deepwater rice is its low yield[2]. Deepwater rice bred for these areas must possess not only internode elongation ability under deepwater conditions, but must also produce higher yields. To date, we have identified three QTLs on chromosomes 1, 3 and 12 that regulate the deepwater response[9,10]. To evaluate the deepwater responses of these QTLs, NILs and QTL pyramiding lines[21,22] were produced. Four QTL pyramiding lines were produced, possessing two QTLs on chromosomes 1 and 3 (NIL-1+3), 3 and 12 (NIL-3+12), 1 and 12 (NIL-1+12), and three QTLs on chromosomes 1, 3 and 12 (NIL-1+3+12) in a T65 genetic background (Fig. 4c, d). In particular, the QTL pyramiding line NIL-1+3+12 showed almost the same deepwater response as C9285. These results indicate that introducing these three QTLs into non-deepwater rice enables it to behave like deepwater rice. Therefore, the QTL pyramiding strategy is a powerful and efficient tool for breeding rice varieties for flood-prone areas. Recently, genes that regulate yield have been identified[21,23–26]. The combination of deepwater and yield QTLs has the potential to help breeding in these areas.

In this study, we showed that *SK* genes encoding ERFs trigger internode elongation in deepwater rice. Another water tolerance gene, *SUB1A*, which encodes an ERF (located on chromosome 9), was previously identified in rice[14,15]. *SUB1A* restricts plant elongation during flash floods at the seedling stage. In general, flash floods only continue for a few weeks. Plants carrying *SUB1A* show stunting and can survive in water for a few weeks. They avoid the energy consumption associated with plant elongation, and plant growth restarts after the flood recedes. We tested for the *SUB1A* allele and found that it is
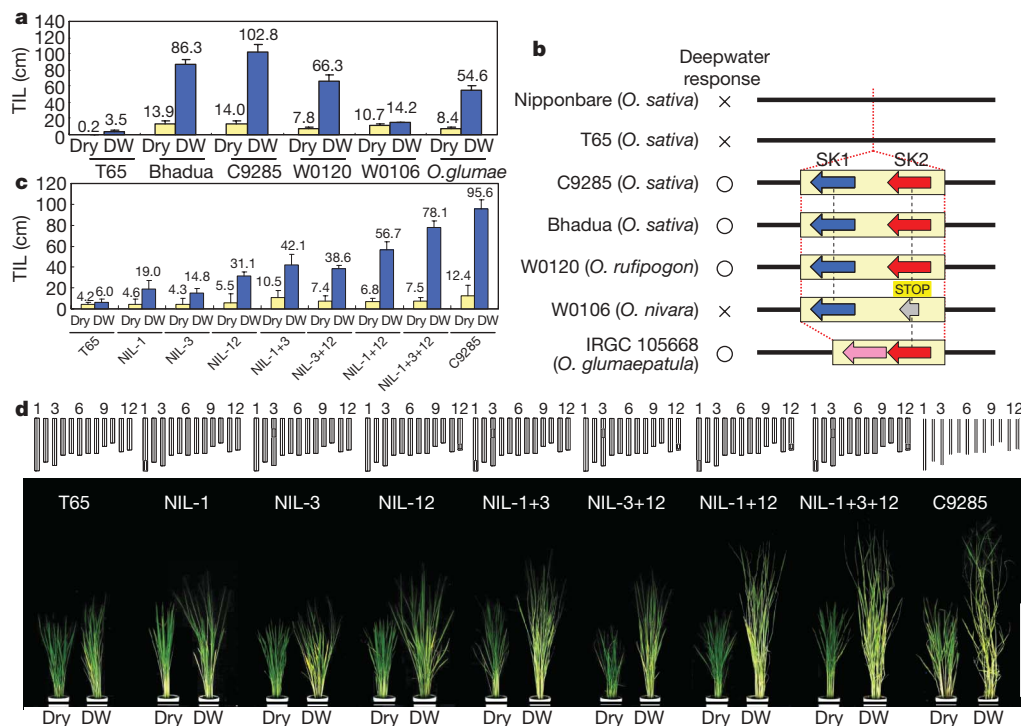
Figure 4 | *SK* genes in wild rice species and QTL pyramiding.
a, Quantitative evaluation of deepwater response in six rice lines. *O. glumae*, *O. glumaepatula*; mean ± s.d., *n* = 5. **b**, Genome structures of the *SK* regions in cultivars and wild rice species. The region corresponding to *SK* genes in each plant are shown in yellow boxes. Blue, red and pink arrows indicate *SK1*, *SK2* and *SK2*-like gene, respectively. **c**, Quantitative evaluation of deepwater response in NILs and pyramiding lines. Mean + s.d., *n* = 5. **d**, QTL pyramiding for rice breeding. Scale bar, 1 m. Ten-leaf-stage plants were submerged for two weeks (**a**, **c**).

absent in both C9285 and T65 (Supplementary Fig. 11, Supplementary Information). *SUB1A* negatively regulates GA responses via restriction of SLR1 and SLRL1 degradation[27]. In contrast, *SK1* and *SK2* may stimulate GA. Both *SK* genes and *SUB1A* encode ERFs and are connected to GA, but they have opposing functions in regulating plant growth in response to flooding. Further investigation of the functional diversity of *SK1*, *SK2* and *SUB1A* is required.

We have identified the genes that allow rice to use natural variation to adapt to deepwater conditions. Loss-of-function mutants are commonly used to identify gene function, with cultivars as the wild type. However, as shown in this study, agriculturally important genes or unique characteristics, such as environmental adaptability, may have been lost in cultivars. In such cases, the wild rice species can be considered wild types and the cultivars as mutants. The study of wild species provides an opportunity to find agriculturally important genes that can be used to develop new varieties and improve existing cultivars.

METHODS SUMMARY

Three deepwater rice lines, W0120, C9285 and Bhadua, two non-deepwater rice lines, W0106 and T65, and one *O. glumaepatula* strain were used in this study. Plants were grown for an appropriate period and subjected to deepwater treatment in water up to 70% of the plant height, and then the total internode elongation length from base to the topmost node was measured. Molecular cloning of *SK1* and *SK2* was performed by a positional cloning and gain-of-function method. Gene expression analyses were performed by RT–PCR using gene specific primers. Microscopic observations were conducted using a GFP to detect the subcellular localization of SK proteins. Details of the above experiments and all other experiments performed in this paper are described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 April; accepted 30 June 2009.

1. Vergara, B. S., Jackson, B. & De Datta, S. K. in *Climate and Rice* (eds IRRI) 301–319 (IRRI, Los Baños, 1976).
2. Catling, D. *Rice in Deepwater* (Macmillan, London, 1992).
3. Kende, H., Van der Knaap, E. & Cho, H.-T. Deepwater rice: a model plant to study stem elongation. *Plant Physiol.* **118**, 1105–1110 (1998).
4. Métraux, J.-P. & Kende, H. The role of ethylene in the growth response of submerged deep water rice. *Plant Physiol.* **72**, 441–446 (1983).
5. Raskin, I. & Kende, H. Role of gibberellin in the growth response of submerged deep water rice. *Plant Physiol.* **76**, 947–950 (1984).
6. Hoffmann-Benning, S. & Kende, H. On the role of abscisic acid and gibberellin in the regulation of growth in rice. *Plant Physiol.* **99**, 1156–1161 (1992).
7. Azuma, T. *et al.* Involvement of the decrease in levels of abscisic acid in the internodal elongation of submerged floating rice. *J. Plant Physiol.* **14**, 323–328 (1995).
8. Bailey-Serres, J. & Voesenek, L. A. C. J. Flooding stress: acclimations and genetic diversity. *Annu. Rev. Plant Biol.* **59**, 313–339 (2008).
9. Hattori, Y. *et al.* A major QTL confers rapid internode elongation in response to water rise in deepwater rice. *Breed. Sci.* **57**, 305–314 (2007).
10. Hattori, Y. *et al.* Mapping of three QTLs that regulate internode elongation in deepwater rice. *Breed. Sci.* **58**, 39–46 (2008).
11. Nemoto, K. *et al.* Inheritance of early elongation ability in floating rice revealed by diallel and QTL analyses. *Theor. Appl. Genet.* **109**, 42–47 (2004).
12. Kawano, R. *et al.* Mapping of QTLs for floating ability in rice. *Breed. Sci.* **58**, 47–53 (2008).
13. Nakano, T., Suzuki, K., Fujimura, T. & Shinshi, H. Genome-wide analysis of the ERF gene family in *Arabidopsis* and Rice. *Plant Physiol.* **140**, 411–432 (2006).
14. Xu, K. *et al. Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**, 705–708 (2006).
15. Fukao, T., Xu, K., Ronald, P. C. & Bailey-Serres, J. A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell* **18**, 2021–2034 (2006).
16. Solano, R., Stepanova, A., Chao, Q. & Ecker, J. R. Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* **12**, 3703–3714 (1998).
17. Jackson, M. B. Ethylene and responses of plants to soil waterlogging and submergence. *Annu. Rev. Plant Physiol.* **36**, 145–174 (1985).
18. Hooley, R. Gibberellins: perception, transduction and responses. *Plant Mol. Biol.* **26**, 1529–1555 (1994).
19. Cheng, C., Tsuchimoto, S., Ohtsubo, H. & Ohtsubo, E. Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. *Genes Genet. Syst.* **77**, 323–334 (2002).
20. Vaughan, D. A., Lu, B.-R. & Tomooka, N. The evolving story of rice evolution. *Plant Sci.* **174**, 394–408 (2008).
21. Ashikari, M. *et al.* Cytokinin oxidase regulates rice grain production. *Science* **309**, 741–745 (2005).

22. Ashikari, M. & Matsuoka, M. Identification, isolation and pyramiding of quantitative trait loci for rice breeding. *Trends Plant Sci.* **11**, 344–350 (2006).
23. Fan, C. *et al.* GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–1171 (2006).
24. Song, X.-J. *et al.* A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nature Genet.* **39**, 623–630 (2007).
25. Shomura, A. *et al.* Deletion in a gene associated with grain size increased yields during rice domestication. *Nature Genet.* **40**, 1023–1028 (2008).
26. Huang, X. *et al.* Natural variation at the *DEP1* locus enhances grain yield in rice. *Nature Genet.* **41**, 494–497 (2009).
27. Fukao, T. & Bailey-Serres, J. Submergence tolerance conferred by *Sub1A* is mediated by SLR1 and SLRL1 restriction of gibberellin responses in rice. *Proc. Natl Acad. Sci. USA* **105**, 16814–16819 (2008).

Author Contributions M.A. conceived the project and designed the experiments. Y.H. identified the genes and Y.H., K.N., S.F., X.-J.S. and R.K. performed molecular characterization of the genes. H.S. and H.M. surveyed the hormone contents. J.W. and T.M. performed BAC clone analysis. A.Y., H.K. and M.M. provided advice regarding the experiments. M.A. and Y.H. wrote the manuscript.

Author Information The DDBJ accession numbers for *SNORKEL1* and *SNORKEL2* are as follows (rice variety, accession numbers): C9285, AB510478 and AB510479; Bhadua, AB510480 and AB510481; *O. rufipogon* (W0120), AB510482 and AB510483; and in *O. nivara* (W0106), AB510484 and AB510485. *SNORKEL2* and *SNORKEL2*-like genes in *O. glumaepatula* (IRGC105668) are AB510486 and AB510487. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.A. (ashi@agr.nagoya-u.ac.jp).

METHODS

Plant materials. The Indian wild rice species W0120 (*O. rufipogon*; perennial), which tolerates deepwater conditions[28], and the deepwater rice cultivars C9285 and Bhadua (*O. sativa*, ssp. *indica*) from Bangladesh were used as deepwater rice plants. An Indian wild rice species, W0106 (*O. nivara*; also called annual type *O. rufipogon*), and the cultivar Taichung65 (T65; *O. sativa* ssp. *japonica*) were used as rice plants lacking deepwater characteristics (non-deepwater rice plants) and as crossing material. The South American wild rice species *O. glumaepatula* (IRGC105668) and other *O. nivara* strains (IRGC105319, IRGC105703, IRGC105715 and IRGC106052) were also used in this study. The lines C9285, W0120, W0106 and Bhadua were kindly provided by the National Institute of Genetics of Japan, and IRGC105668, IRGC105319, IRGC105703, IRGC105715 and IRGC106052 was obtained from the International Rice Research Institute. T65 was maintained at Nagoya University.

Growth conditions and phenotypic evaluation. Plants were germinated and transplanted into small pots. During growth, the water level was beneath the soil surface (dry conditions, described as 'Dry' in figure). Plants were submerged in water up to 70% of the plant height. To evaluate plant growth as shown in Figs 1b, c, 3b, 4a, c and Supplementary Fig. 5, plants at the ten-leaf-stage were submerged for two weeks (Figs 1b, c, 4a, c) or one week with each treatment (Fig. 3b, Supplementary Fig. 5) before determining the TIL (total internode elongation length)[9]. For expression analysis of *SK*s and other genes in Figs 1f, g, 2c and Supplementary Fig. 4, ten-leaf-stage plants were submerged for indicated times (Figs 1f, 2c, Supplementary Fig. 4) and 12 h (Fig. 1g), and in Fig. 2e–g, six-leaf-stage plants were subjected to each treatment for one week (Fig. 2e, f) or for indicated times (Fig. 2g). For measurement of phytohormones contents in Fig. 3a and Supplementary Table 4, six-leaf-stage plants were submerged for the indicated time and the nodes were sampled.

Preparation of BAC libraries. BAC libraries were constructed from young leaves of C9285 (*O. sativa*), Bhadua (*O. sativa*), W0120 (*O. rufipogon*) and IRGC105668 (*O. glumaepatula*) by conventional methods through partial DNA digestion with HindIII, size fractionation of high molecular weight DNA by pulsed-field gel electrophoresis (CHEF, Bio-Rad Laboratories), vector ligation (pIndigo BAC-5; Epicentre Biotechnologies), and transformation into *Escherichia coli* (DH10B). Positive BAC clones were screened by PCR from the pooled DNA or by Southern hybridization from high-density BAC filters of each library that contained clones with insert DNA of sufficiently large size for at least sixfold genome coverage. A BAC clone of T65 (GN_42I09) was selected from BAC libraries and a BAC clone of W0106 was purchased from the Arizona Genomic Institute of the University of Arizona.

Production of NILs and pyramiding lines. The deepwater rice cultivar C9285 was crossed with the non-deepwater cultivar T65. NILs (NIL-1, NIL-3 and NIL-12: BC_4F_2) were developed by repeated backcrossing with T65. The pyramiding lines NIL-1+3, NIL-3+12, NIL-1+12 and NIL-1+3+12 were obtained by crossing with NILs.

Molecular cloning of *SK1* and *SK2*. A total of 16,000 progeny of heterozygote NIL-12 from T65/C9285 (BC_4F_2) and 12,000 progeny of heterozygote NIL-12B from T65/Bhadua (BC_4F_2) were used for positional cloning of *SK*s. Phenotypic evaluation of *SK*s under deepwater conditions was confirmed using BC_4F_3 (see Supplementary Information for details).

Full-length genomic DNA of *SK*s including the promoter region was inserted into the binary vector pYLTAC7. These DNA fragments were introduced into T65 by *Agrobacterium tumefaciens*-mediated transformation[29]. The empty vector was also introduced into T65 as a control.

To obtain the full-length sequences of *SK1* and *SK2*, 5′ and 3′ RACE was performed using 5′-Full RACE Core Set (TaKaRa) and 3′-Full RACE Core Set (TaKaRa) in accordance with the manufacturer's instructions.

RNA isolation and expression analysis. Total RNA was prepared as described previously[30]. First-strand cDNA was synthesized from 2 μg of total RNA using Omniscript Reverse Transcription kit (Qiagen). RT–PCR was performed as described previously[31] using gene-specific primers (Supplementary Table 1).

Plasmid construction and plant transformation. To construct plants over-expressing *SK*s, their cDNA was amplified and inserted into the pBI101 binary vector containing an *OsAct1* promoter. Plasmids were transformed into T65 as described previously[32]. Transgenic plants were selected by hygromycin resistance.

To construct GFP fusion proteins, *SK*s cDNA was amplified by PCR using gene-specific primers and fragments were inserted into the destination vector pUC119 to generate fusion constructs with GFP at the C-terminal of *SK*s under the control of the *OsAct1* promoter.

Microscopic observations. Gold particles were coated with the *pAct::SK1-GFP* and *pAct::SK2-GFP* fusion constructs and used to bombard onion epidermis with a PDS-1000/He biolistic system (Bio-Rad). The onion epidermis was incubated in the dark at 22 °C. After 24 h, the cell layers were soaked in 2 μg μl^{-1} 4′,6-diamidino-2-phenylindole dihydrochloride *n*-hydrate (DAPI; Dojindo) solution for visualization of the nucleus and analysis of the nuclear localization of pAct::SK1-GFP and pAct::SK2-GFP derivatives. The stained samples were observed with a confocal microscanning laser microscope (FV500; Olympus).

Transactivation activity assay. The GAL4-based Matchmaker Two-Hybrid System 3 (Clontech) was used for the transactivation activity assay. To construct pGBKT7-BD-SKs-Full, pGBKT7-BD-SKs-NT, pGBKT7-BD-SKs-AP2/ERF and pGBKT7-BD-SKs-CT, the full-length coding sequence, N-terminal, C-terminal, or AP2/ERF domain regions alone of *SK*s were amplified using PCR and confirmed by sequencing. The validated PCR products were inserted into pGBKT7 and fused with the GAL4-binding domain after digestion with EcoRI and PstI. All constructs were transformed into the yeast strain AH109. Each yeast liquid culture was serially diluted to $A_{600} = 0.6$, and 2 μl of each dilution was inoculated onto tryptophan- and histidine-negative synthetic dropout medium.

Quantification of phytohormones. Phytohormones (auxins, cytokinins, gibberellins and abscisic acid) were extracted from samples of about 100 mg (fresh weight) of the rice nodes. The phytohormones were quantified as described[33] using an LC/MS system (UPLC/Quattro Premier XE; Waters) with an ODS column (Acquity-UPLC BEH-C$_{18}$, 1.7 μm, 2.1 × 100 mm; Waters). A special device was used to measure ethylene (Supplementary Fig. 10). The plants were submerged in water for indicated times and then immediately transferred into a vacuum container filled with saturated NaCl solution. To collect the internal gas, the plant body was subjected to vacuum for 90 s, and emitted gas was trapped in the graduated test tube. After release from vacuum, the tube was sealed with rubber stopper in saturated NaCl solution. One millilitre of headspace gas in the tube was injected into a gas chromatograph (model GC353, GL Sciences Inc.) fitted with a flame-ionization detector and PoraBond Q capillary column (Varian Inc.).

Phytohormones and inhibitor treatment. For analysis of phytohormone responses, ten-leaf-stage plants were transferred to the closed chamber with 10 p.p.m. ethylene or water containing 10 μM GA, 10 nM BR, 20 μM IAA, 1 μM CK, or 100 μM ABA. Uniconazole (1 μM) and 1-MCP (10 p.p.m.) were used as a GA biosynthesis inhibitor and ethylene perception inhibitor. Plants were pretreated with both inhibitors for 3 days before the experiments.

Electrophoresis mobility shift assays (EMSAs). Full-length *OsEIL1b*[34] was amplified by PCR and inserted into pET-32a(+) vector (Novagen) with the His-tag coding sequence and overexpressed in Rosetta (DE3) (Novagen) to generate the OsEIL1b. The recombinant proteins were purified with a His-tag affinity column (Bio-Rad) and Superdex-200 gel filtration chromatography (GE Healthcare). The promoter fragment probes of *SK*s (Supplementary Table 3) were labelled with [^{32}P]dATP. The DNA binding reaction was allowed to proceed for 30 min at 4 °C in 20 μl of binding buffer (12.5 mM Tris-HCl, 60 mM NaCl, 0.25 mM DTT, 12.5% glycerin, 1 mM EDTA, 0.05% NP-40, and 2 μg of poly(dI-dC)·poly(dI-dC)) containing 0.5 ng of ^{32}P-labelled oligonucleotide probe and recombinant protein, and then electrophoresed through a 13% poly-acrylamide gel in 0.25 × Tris-borate-EDTA buffer. Competition experiments were performed by adding unlabelled competitor oligonucleotides to the binding reaction with the subsequent addition of labelled oligonucleotides. Mutated forms of OsEIL1b binding site were obtained by the overlapping primer method (see Supplementary Table 3) and used in competition experiments.

28. Morishima, H., Hinata, K. & Oka, H. I. Floating ability and drought resistance in wild and cultivated species of rice. *Ind. J. Genet. Plant Breed.* **22**, 1–11 (1962).

29. Hood, E. E., Helmer, G. L., Fraley, R. T. & Chilton, M.-D. The hypervirulence of *Agrobacterium tumefaciens* A281 is encoded in a region of pTiBo542 outside of T-DNA. *J. Bacteriol.* **168**, 1291–1301 (1986).

30. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1989).

31. Kaneko, M. *et al.* Where do gibberellin biosynthesis and gibberellin signaling occur in rice plants? *Plant J.* **35**, 104–115 (2003).

32. Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).

33. Hirano, K. *et al.* Comprehensive transcriptome analysis of phytohormone biosynthesis and signaling genes in microspore/pollen and tapetum of rice. *Plant Cell Physiol.* **49**, 1429–1450 (2008).

34. Mao, C., Wang, S., Jia, Q. & Wu, P. *OsEIL1*, a rice homolog of the *Arabidopsis EIN3* regulates the ethylene response as a positive component. *Plant Mol. Biol.* **61**, 141–152 (2006).

LETTERS

Characterization of two classes of small molecule inhibitors of Arp2/3 complex

B. J. Nolen[1,2,3]†*, N. Tomasevic[4]†*, A. Russell[4], D. W. Pierce[4]†, Z. Jia[4], C. D. McCormick[3], J. Hartman[4], R. Sakowicz[4]† & T. D. Pollard[1,2,3]

Polymerization of actin filaments directed by the actin-related protein (Arp)2/3 complex supports many types of cellular movements[1]. However, questions remain regarding the relative contributions of Arp2/3 complex versus other mechanisms of actin filament nucleation to processes such as path finding by neuronal growth cones; this is because of the lack of simple methods to inhibit Arp2/3 complex reversibly in living cells. Here we describe two classes of small molecules that bind to different sites on the Arp2/3 complex and inhibit its ability to nucleate actin filaments. CK-0944636 binds between Arp2 and Arp3, where it appears to block movement of Arp2 and Arp3 into their active conformation. CK-0993548 inserts into the hydrophobic core of Arp3 and alters its conformation. Both classes of compounds inhibit formation of actin filament comet tails by *Listeria* and podosomes by monocytes. Two inhibitors with different mechanisms of action provide a powerful approach for studying the Arp2/3 complex in living cells.
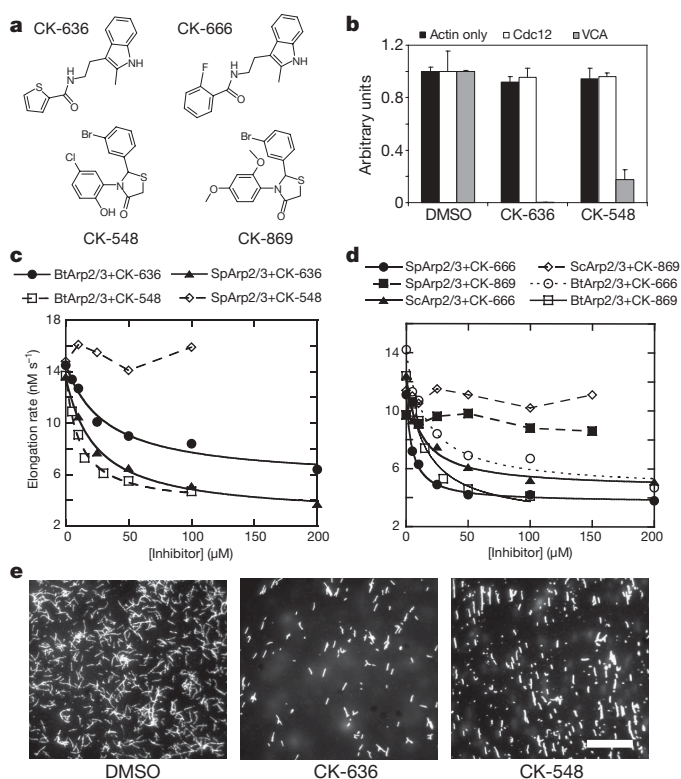
We used fluorescence assays to screen a library of >400,000 small molecules for inhibitors of actin polymerization stimulated by human (*Homo sapiens*, Hs) Arp2/3 complex and either full-length human Wiskott–Aldrich syndrome protein (HsWASP) with acrylodan-actin or HsWASP residues 105–502 with pyrenyl-actin. These screens each identified two inhibitors of human and bovine (*Bos taurus*, Bt) Arp2/3 complex, namely CK-0944636 (abbreviated CK-636) and CK-0993548 (abbreviated CK-548) (Fig. 1a). These compounds inhibited BtArp2/3 complex with IC_{50} (half-maximal inhibitory concentration) values of 32 μM for CK-636 and 11 μM for CK-548 (Fig. 1c). CK-636 inhibited actin polymerization stimulated by fission yeast (*Schizosaccharomyces pombe*, Sp) Arp2/3 complex ($IC_{50} = 24$ μM), but 100 μM CK-548 did not (Fig. 1c, Supplementary Table 1). Fluorescence microscopy of the products of these reactions stained with Alexa 488 phalloidin showed branched actin filaments in controls (Fig. 1e, left panel). Samples with 100 μM CK-636 contained fewer branched filaments (Fig. 1e, centre panel), while samples with 100 μM CK-548 contained only unbranched filaments (Fig. 1e, right panel). We tested a number of compounds structurally related to CK-548 or CK-636 that had no effect on actin polymerization at concentrations up to 200 μM and are useful as

Figure 1 | Two classes of small molecules inhibit nucleation of actin filaments by Arp2/3 complex. a, Structures of CK-636, CK-548, CK-666 and CK-869. **b,** Inhibition of HsArp2/3 complex by CK-636 and CK-548. The time course of actin polymerization was monitored by the fluorescence increase of pyrenyl-actin. Conditions: 20 μM compound or DMSO, 2.5 μM 15% pyrenyl-actin alone or with 100 nM Cdc12(FH2) or 6 nM HsArp2/3 complex and 300 nM WASP-VCA. The maximum polymerization rate is expressed in arbitrary units. Error bars, s.d.; $n = 4$). **c,** Inhibition of Bt and Sp Arp2/3 complexes by CK-636 and CK-548. The time course of polymerization was measured as in **b**. Conditions: 4 μM 15% pyrenyl-actin, 5 nM SpArp2/3 complex, and 1 μM N-WASP-VCA, or 3 μM 30% pyrenyl-actin, 5 nM BtArp2/3 complex and 1 μM N-WASP-VCA. CK-548 was insoluble at 200 μM under the conditions used for this assay. The maximum polymerization rate of actin alone under these conditions was 4.6 nM s^{-1}. **d,** Effect of CK-666 and CK-869 on the polymerization of actin with bovine and yeast Arp2/3 complexes. Conditions as in **c** with either 3 μM 30% pyrenyl-actin and 5 nM BtArp2/3 complex or 4 μM 20% pyrenyl-actin plus 20 nM SpArp2/3 complex or 5 nM ScArp2/3 complex. Both compounds reduced the maximum polymerization rate of samples with BtArp2/3 complex to the basal rate without Arp2/3 complex, but CK869 did not inhibit either yeast Arp2/3 complex. **e,** Fluorescence micrographs of the products of actin polymerization assays stained with Alexa 488-phalloidin. Actin (3.6 μM) was polymerized with 6 nM HsArp2/3 complex, 300 nM Cdc42, 100 μM CK-636 and 100 nM of a WASP construct lacking only its N-terminal Ena/Vasp homology domain, WASP-105–502. Scale bar, 20 μm.

[1]Department of Molecular Cellular and Developmental Biology, [2]Department of Cell Biology, [3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. [4]Cytokinetics, Inc., South San Francisco, California 94080, USA. †Present addresses: Department of Chemistry and the Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA (B.J.N.); Kalobios Pharmaceuticals, Inc., South San Francisco, California 94080, USA (N.T.); Five Prime Therapeutics, San Francisco, California 94158, USA (D.W.P.); Gilead Sciences, Inc., Foster City, California 94404, USA (R.S.).
*These authors contributed equally to this work.

controls for experiments with cells (Fig. 2g). Supplementary Table 2 lists one inactive compound from each class.

The following control experiments showed that both compounds interact with Arp2/3 complex rather than actin or WASP. At 20 μM, neither compound inhibited polymerization of actin alone or polymerization stimulated by the FH2 domain of formin Cdc12p (Cdc12(FH2); Fig. 1b), so neither compound interacts directly with actin. As actin can polymerize on its own, neither compound eliminated actin polymerization in mixtures with Arp2/3 complex and WASP or a carboxy-terminal construct of Neural Wiskott–Aldrich syndrome protein (N-WASP-VCA) (Fig. 1c, d). Both compounds inhibited polymerization stimulated by HsArp2/3 complex and ActA, an Arp2/3 complex activating factor from *Listeria monocytogenes*[2], suggesting that they do not act on WASP (Supplementary Fig. 1). A fluorescence anisotropy assay showed that neither compound had a large effect on Arp2/3 complex binding N-WASP-VCA (Supplementary Fig. 2). The affinity of rhodamine-N-WASP-VCA for BtArp2/3 complex was the same with 50 μM CK-636 ($K_d = 470 \pm 50$ nM) as controls ($K_d = 510 \pm 30$ nM), but 50 μM CK-548 decreased the affinity approximately twofold ($K_d = 1.0 \pm 0.2$ μM).

We used a *Listeria* motility assay to determine whether CK-636 and CK-548 can inhibit actin polymerization mediated by Arp2/3 complex in live cells. Both compounds reduced the formation of actin filament comet tails by *Listeria* in infected SKOV3 cells (Fig. 2a–c). The concentration dependence of this inhibition gave IC_{50} values of 22 μM for CK-636 and 31 μM for CK-548. We used the actin polymerization and *Listeria* comet tail assays to search for more potent inhibitors related to CK-636 and CK-548.

Compound CK-0944666 (abbreviated CK-666) has a fluorobenzene rather than the thiophene ring of CK-636 (Fig. 1a). CK-666 was a better inhibitor of actin polymerization with either BtArp2/3 complex ($IC_{50} = 32$ μM for CK-636 verses 17 μM for CK-666) or SpArp2/3 complex ($IC_{50} = 24$ μM for CK-636 versus 5 μM for CK-666) (Fig. 1d, Supplementary Table 1). Both CK-636 and CK-666 had IC_{50} values of 4 μM for inhibiting HsArp2/3 complex. CK-666 reduced actin polymerization around intracellular *Listeria* to background levels (Fig. 2d, e and g) at lower concentrations than CK-636 (Fig. 2c). Actin filament haloes and comet tails reformed when the compounds were removed after one hour of treatment (Fig. 2f, g). Inactive control compound CK-689 had no effect on comet tail formation (Fig. 2g).

CK-0157869 (abbreviated CK-869) has methoxy groups in the *ortho* and *para* positions replacing the *ortho*-hydroxy and *meta*-chlorine substituents of CK-548 (Fig. 1a). CK-869 inhibited actin polymerization with BtArp2/3 complex similarly to CK-548 ($IC_{50} = 11$ μM for both compounds), but CK869 inhibited comet tail formation by *Listeria* more effectively ($IC_{50} = 7$ μM for CK-869 versus 31 μM for CK-548). Like CK-548, CK-869 did not inhibit either budding or fission yeast Arp2/3 complexes (Fig. 1d).

Both CK-548 and CK-636 also inhibited the formation of podosomes by the THP-1 monocyte cell line (Fig. 2h–k). Podosomes are adhesive structures that depend on Arp2/3 complex, WASP and actin polymerization[3]. Podosomes also depend on microtubules[4], but neither CK-548 or CK-636 at 100 μM disrupted the microtubule network in THP-1 cells (Supplementary Fig. 3). Neither class of compound shows irreversible or off-target morphological effects on cells, such as apoptosis, at concentrations up to 100 μM over 24 h. Our most
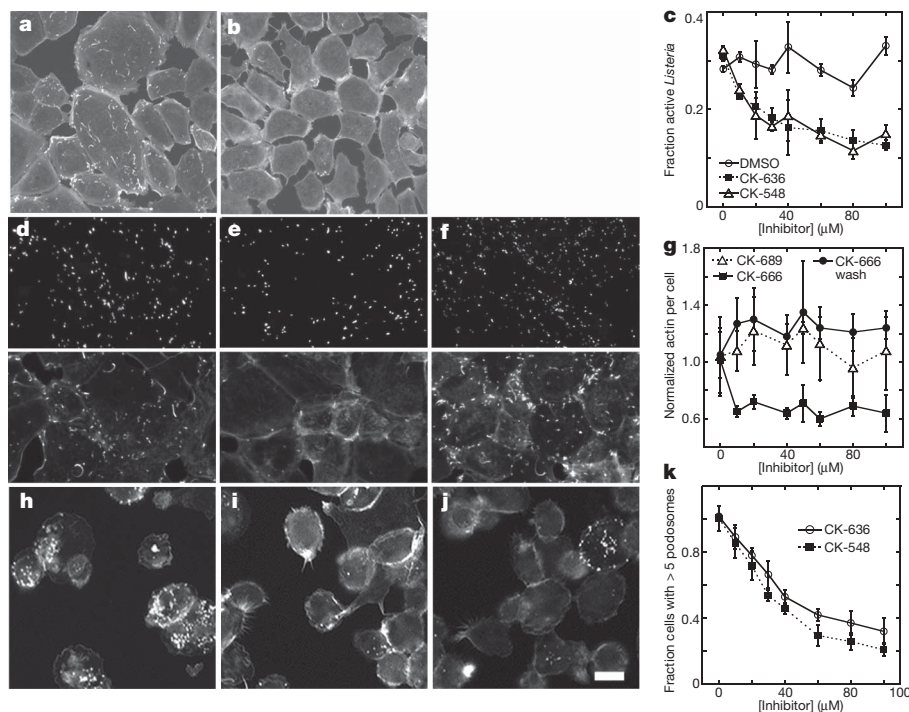
Figure 2 | Inhibition of actin assembly in live cells by CK-548, CK-636 and CK-666. a–g, Formation of actin filament comet tails by *Listeria* infecting SKOV3 cells. **a, b,** Fluorescence micrographs of fixed cells stained with rhodamine phalloidin. **a,** Cells incubated at 37 °C for 90 min with 0.1% DMSO had *Listeria* comet tails. **b,** Cells incubated at 37 °C for 90 min with 100 μM CK-548 in 0.1% DMSO had no actin comet tails. **c,** Dependence of the fraction of *Listeria* with comet tails on the concentrations of CK-636 and CK-548. Error bars, s.d.; *n* = 3. **d–g,** Effects of CK-666 on actin fluorescence around *Listeria* in SKOV3 cells. Infected cells were treated with 40 μM CK-666 for 60 min followed by a 60 min washout. The pairs of fluorescence micrographs show anti-*Listeria* fluorescence (top) and Alexa 488 phalloidin fluorescence (bottom). **d,** Control without CK-666 for 60 min. **e,** CK-666 for 60 min. **f,** CK-666 for 60 min followed by 60 min washout. **g,** Dependence of

the mean actin fluorescence around each *Listeria* cell on the concentrations of CK-666 (active) and CK-689 (inactive) for 60 min. Fluorescence recovered when CK-666 was washed out for 60 min. Error bars, s.d. of the mean fluorescence values from four separate experiments, each with around 5,000 *Listeria* per condition. A fluorescence value of 0.6 corresponds to background actin fluorescence. **h–k,** Effect of compounds on podosome formation by THP-1 derived monocyte cells. Adhered cells were treated with DMSO or compounds for 15 min, then fixed and stained with Alexa 568 phalloidin. **h,** 0.1% DMSO control. **i,** 100 μM CK-636 and **j,** 100 μM CK-548 reduce formation of podosomes. **k,** Fraction of cells with >5 podosomes versus inhibitor concentration, normalized to DMSO control. Error bars, s.d.; *n* = 3. Scale bar, 20 μm.

potent compound, CK-666, had no effect on the mitotic index of SKOV cells at concentrations up to 80 μM, while inhibiting actin assembly around *Listeria* completely at 10 μM. Fifty micromolar CK-666 or CK-869 each altered the morphology and slightly slowed but did not stop the gliding motility of fish keratocytes (data not shown), as expected, as ~10% of Arp2/3 complex (~500 nM) would still be active. Thus *Listeria* comet tails and monocyte podosomes are more sensitive than lamellar motility to inhibition of Arp2/3 complex.

We determined X-ray crystal structures of BtArp2/3 complex with either CK-636 or CK-548 bound, using molecular replacement with the apo-form BtArp2/3 complex as the search model (PDB accession 1K8K)[5]. Structures of Arp2/3 complex co-crystallized with the compounds and with compounds soaked into the crystals were nearly identical, but the resolution of soaked crystals was higher (2.70 Å for CK-636 and 2.85 Å for CK-548, Supplementary Table 3). As in previous crystal structures of BtArp2/3 complex[5–7], none of these structures had appreciable electron density for subdomains 1 and 2 of Arp2, which appears only after chemical crosslinking[7].

After one round of rigid body refinement, strong $F_o - F_c$ electron density was present at the interface between Arp2 and Arp3 in maps calculated with data from crystals soaked in 1 mM CK-636 (Fig. 3). This density improved on further refinement (Supplementary Fig. 4a), and CK-636 was modelled into the density when the R_f was 29.2%. CK-636 binds in a pocket between subdomain 4 of Arp2 and subdomain 1 of Arp3. The contact surface is mainly hydrophobic and buries 161 Å² of the accessible surface of Arp2 and 79 Å² of Arp3. CK-636 forms hydrogen bonds with two residues in the pocket. The side chain of Asp 248 forms a hydrogen bond with the nitrogen in the indole ring of the inhibitor, and the backbone amide of Ala 203 forms a hydrogen bond with the carbonyl oxygen on the linker between the two ring systems. Both *trans* and *cis* conformations of the amide functional group of CK-636 fit into the electron density with the hydrogen bonds described above preserved in both conformations.
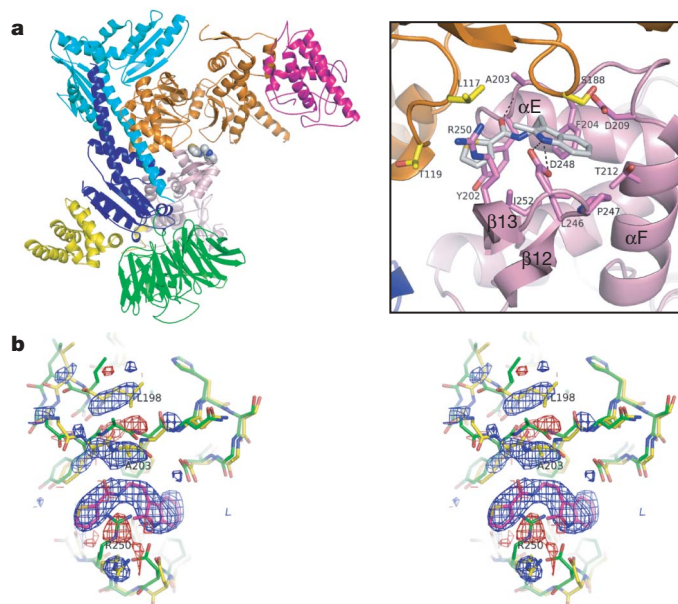
Figure 3 | Crystal structure of BtArp2/3 complex with bound CK-636. Colour code: Arp3, orange; Arp2, pink; actin-related protein complex, subunit 1 (ARPC1), green; ARPC2, cyan; ARPC3, magenta, ARPC4, blue; ARPC5, yellow. **a**, Ribbon diagram with (right) a detail of binding pocket. **b**, Stereo diagram of an $F_o - F_c$ electron density map, showing changes caused by CK-636. Positive (blue) and negative (red) difference densities indicate the position of CK-636 and small conformational changes of the protein. The apo-structure (PDB accession 1K8K) is shown in green stick representation, and the final structure in yellow. The density is contoured at 3.5σ, and was generated using structure factors calculated after one round of rigid body refinement and the data for the CK-636-soaked crystal.

Because *trans*-amides are generally 1–5 kcal mol^{-1} more stable than *cis*-amides in solution[8], and the *cis* conformation is rarely observed in small molecules, we modelled CK-636 in the *trans* conformation.

A difference electron density map shows that CK-636 causes minor conformational changes in the complex (Fig. 3b). Arp2 Arg 250 from β-strand 13 rotates out of the pocket and adopts an extended conformation. Arp2 Ala 203 from the loop (residues 199–208) connecting α-helices αE and αF moves towards the pocket, causing the side chain of Leu 198 to adopt a new rotamer. These changes and small changes in the backbone of residues surrounding the pocket result in an overall r.m.s. deviation of 0.63 Å for an overlay of subdomains 3 and 4 (191 Cα atoms) of Arp2 with and without inhibitor bound.

We modelled CK-666 into the CK-636 binding pocket with the aromatic fluorine pointed towards a concave surface in the back wall of the pocket formed by Tyr 202 in Arp2 and residues 118 to 120 in Arp3 (Supplementary Fig. 6). One hundred steps of conjugate gradient minimization in CNS[9] showed that CK-666 is stable in this conformation and does not clash with residues in the pocket. The fluorine atom provides additional van der Waals interactions with the back wall of the pocket, and the benzene ring completely fills the hydrophobic pocket created by Ile 252, Tyr 202 and the aliphatic portion of Thr 119. These interactions may allow CK-666 to bind more tightly than CK-636.

During formation of a branch Arp2 moves 31 Å relative to Arp3 from its position in inactive Arp2/3 complex[10], so the location of the CK-636 binding pocket between Arp2 and Arp3 suggests that CK-636 and CK-666 lock Arp2/3 complex in an inactive conformation. CK-636 at a concentration of 100 μM did not interfere with N-WASP-VCA binding to Arp2/3 complex (Supplementary Fig. 2), but it prevented WASP-VCA binding from increasing the fluorescence of Arp2/3 complex loaded with etheno-ATP (Supplementary Fig. 5). Thus CK-636 inhibits a conformational change caused by activator binding. This supports the hypothesis that CK-636 locks the complex in an inactive conformation.

The residues contributed by Arp2 and Arp3 to form the CK-636 binding pocket are conserved across a broad range of species (Supplementary Table 4), so we expect CK-636 and CK-666 will be useful for inhibiting Arp2/3 complex from diverse species. These residues are also conserved in actin, but CK-636 does not inhibit actin polymerization at concentrations up to 200 μM. Only half of the binding site is available on an actin monomer, and the two half-sites are not juxtaposed in actin filaments[11].

CK-548 has a single stereocentre and was used as a racemic mixture in our experiments. The 2S enantiomer binds in a hydrophobic cavity in the core of the Arp3 subunit (Fig. 4a, b and Supplementary Fig. 4b). The isolated 2R enantiomer, which does not fit in the binding pocket, should provide a useful control for *in vivo* experiments. Binding of CK-548 requires a substantial conformational change: the loop (residues 76–85) connecting β6 (73–75) and αB (86–98) in subdomain 1 of Arp3 flips upward 7.2 Å, exposing the binding site in the hydrophobic core of Arp3 (Fig. 4b, c). CK-548 contacts 239 Å² of the inner surface of this cavity, which includes residues Cys 12, Ile 78, Trp 86, Met 89, Met 93 and Leu 112. A hydrogen bond between the side chain amide of Asn 118 and the carbonyl oxygen of CK-548 anchors the inhibitor to the loop between β7 (110–114) and αC (120–132), which forms the bottom lip of the cavity. The ability of CK-548 to bind crystallized Arp2/3 complex suggests that the β6/αB loop can open the hydrophobic binding pocket both in crystals and in solution, even in the absence of inhibitor. The average B-factor for all residues in the loop is 63 Å² in the closed conformation and 82 Å² in the open conformation, suggesting that the loop is more flexible when open.

When modelled into the CK-548 binding site, the *para*-methoxy group of CK-869 projects towards the upper lip of the pocket and is sandwiched by the side chains of Ile 78, Glu 84 and Met 89 (Supplementary Fig. 7). The *ortho*-methoxy group points towards a crease formed at the ends of strands β1 and β7, which could be exploited to design higher affinity inhibitors.

Both CK-548 and CK-869 inhibited human and bovine Arp2/3 complexes, but neither inhibited budding or fission yeast Arp2/3
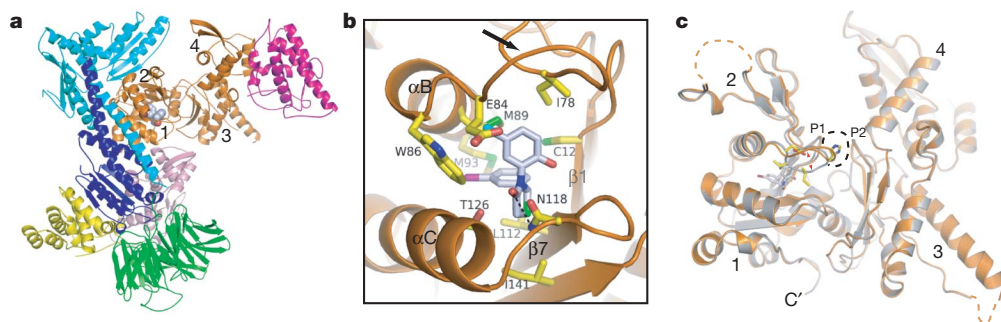
Figure 4 | Crystal structure of BtArp2/3 complex with bound CK-548. The colour coding is the same as in Fig. 3. **a**, Ribbon diagram showing CK-548 bound to Arp3. Subdomains of Arp3 are numbered 1–4. CK-548 is rendered in spheres. **b**, Detail of binding pocket. Black arrow marks the loop (residues 76–85) connecting β6 (73–75) and αB (86–98) in subdomain 1 of Arp3, which flips up to accommodate inhibitor binding. **c**, Ribbon diagrams of Arp3 with bound CK-548 (orange) overlaid onto the apo-BtArp2/3 structure

(PDB 1K8K, grey). Black dashed line indicates the binding pocket for ATP or ADP. P1 and P2 indicate loops that contact the nucleotide phosphates in nucleotide-bound structures. Orange dashed lines indicate disordered regions of the structure. Small red arrowheads indicate alternative conformations of the sensor loop. CK-548 is shown in stick representation with grey carbon atoms and select residues in the sensor loop from the 1K8K structure are shown as sticks with yellow carbon atoms.

complexes (Fig. 1c, d). The residues that contact CK-548 are conserved in Arp2/3 complexes of most mammals, but tryptophan replaces Met 93 at the back of the CK-548 binding pocket in many other species, including budding and fission yeast (Supplementary Table 5). The bent side chain of Met 93 allows CK-548 to bind, but tryptophan at this position blocks binding to non-mammalian Arp3 and actin.

The mechanism of action of CK-548/869 is less apparent than that of CK-636/666. A fluorescence anisotropy assay showed that CK-548 decreased the affinity of rhodamine-labelled N-WASP-VCA for Arp2/3 complex twofold, but this should be inconsequential under the conditions of our assay (Supplementary Fig. 2). The conformation induced by CK-548/869 must interfere with one or more of the reactions along the pathway to branch formation, such as binding of the pointed end of Arp3 to a mother filament[10], nucleotide binding to Arp3, or conformational changes that activate branch formation.

METHODS SUMMARY

Compounds were purchased from Chemdiv: CK-0944636 (catalogue number 8012-5103), CK-0993548 (K205-1650), CK-0944666 (8012-5153) and CK-0157869 (K205-0942). We purified native Arp2/3 complex from human platelets[12], bovine thymus[6], *Schizosaccharomyces pombe*[13] and *Saccharomyces cerevisiae* (Supplementary Methods), actin from chicken skeletal muscle[14], recombinant full length HsWASP from Freestyle 293-FS cells (Invitrogen), HsWASP (residues 105–502), HsWASP-VCA (residues 428–502), Cdc42[12], N-WASP-VCA (residues 428–505) (Supplementary Methods), GST-ActA (residues 36–170) (Supplementary Methods) and *S. pombe* Cdc12p(FH2) (residues 973–1390)[15] from *Escherichia coli*. We used standard assays to measure polymerization of pyrenyl-actin[16] and to visualize actin filaments by fluorescence microscopy[17]. Binding of etheno-ATP to Arp2/3 complex was performed as described previously with slight modifications[18]. We crystallized BtArp2/3 complex[7] with either 0.5 mM CK-548 or 1 mM CK-636 in DMSO or soaked these compounds into crystals for 24 h before freezing in liquid nitrogen. Diffraction data were collected at beamline X29A at Brookhaven National Laboratories. SKOV3 cells were infected with *L. monocytogenes* and fixed with 2% formaldehyde, permeabilized with 0.1% Triton-X in PBS, stained with *Listeria* antibody (US Biologics) and Alexa Fluor 568 phalloidin (Molecular Probes), and imaged by fluorescence microscopy. We used an Isodata threshold on background-subtracted images of *Listeria* to isolate individual bacteria and measure the ratio of colocalized actin to *Listeria* fluorescence. Monocyte THP-1 cells were differentiated in 50 nM phorbol myristate acetate (Sigma-Aldrich Fluka) to form podosomes before treatment with compounds. Black molly keratocytes[19] were observed by time-lapse phase contrast microscopy.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Goley, E. D. & Welch, M. D. The Arp2/3 complex: an actin nucleator comes of age. *Nature Rev. Mol. Cell Biol.* **7**, 713–726 (2006).
2. Welch, M. D., Rosenblatt, J., Skoble, J., Portnoy, D. A. & Mitchison, T. J. Interaction of human Arp2/3 complex and the *Listeria monocytogenes* ActA protein in actin filament nucleation. *Science* **281**, 105–108 (1998).
3. Linder, S. *et al.* The polarization defect of Wiskott–Aldrich syndrome macrophages is linked to dislocation of the Arp2/3 complex. *J. Immunol.* **165**, 221–225 (2000).
4. Gimona, M., Buccione, R., Courtneidge, S. A. & Linder, S. Assembly and biological role of podosomes and invadopodia. *Curr. Opin. Cell Biol.* **20**, 235–241 (2008).
5. Robinson, R. C. *et al.* Crystal structure of Arp2/3 complex. *Science* **294**, 1679–1684 (2001).
6. Nolen, B. J., Littlefield, R. S. & Pollard, T. D. Crystal structures of actin-related protein 2/3 complex with bound ATP or ADP. *Proc. Natl Acad. Sci. USA* **101**, 15627–15632 (2004).
7. Nolen, B. J. & Pollard, T. D. Insights into the influence of nucleotides on actin family proteins from seven structures of Arp2/3 complex. *Mol. Cell* **26**, 449–457 (2007).
8. Jabs, A., Weiss, M. S. & Hilgenfeld, R. Non-proline *cis* peptide bonds in proteins. *J. Mol. Biol.* **286**, 291–305 (1999).
9. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
10. Rouiller, I. *et al.* The structural basis of actin filament branching by the Arp2/3 complex. *J. Cell Biol.* **180**, 887–895 (2008).
11. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347**, 44–49 (1990).
12. Tomasevic, N. *et al.* Differential regulation of WASP and N-WASP by Cdc42, Rac1, Nck, and PI(4,5)P$_2$. *Biochemistry* **46**, 3494–3502 (2007).
13. Nolen, B. J. & Pollard, T. D. Structure and biochemical properties of fission yeast Arp2/3 complex lacking the Arp2 subunit. *J. Biol. Chem.* **283**, 26490–26498 (2008).
14. MacLean-Fletcher, S. & Pollard, T. D. Identification of a factor in conventional muscle actin preparations which inhibits actin filament self-association. *Biochem. Biophys. Res. Commun.* **96**, 18–27 (1980).
15. Kovar, D. R. & Pollard, T. D. Insertional assembly of actin filament barbed ends in association with formins produces piconewton forces. *Proc. Natl Acad. Sci. USA* **101**, 14725–14730 (2004).
16. Cooper, J. A., Walker, S. B. & Pollard, T. D. Pyrene actin: documentation of the validity of a sensitive assay for actin polymerization. *J. Muscle Res. Cell Motil.* **4**, 253–262 (1983).
17. Blanchoin, L. *et al.* Direct observation of dendritic actin filament networks nucleated by Arp2/3 complex and WASP/Scar proteins. *Nature* **404**, 1007–1011 (2000).
18. Dayel, M. J., Holleran, E. A. & Mullins, R. D. Arp2/3 complex requires hydrolyzable ATP for nucleation of new actin filaments. *Proc. Natl Acad. Sci. USA* **98**, 14871–14876 (2001).
19. Lee, J., Ishihara, A., Theriot, J. A. & Jacobson, K. Principles of locomotion for simple-shaped cells. *Nature* **362**, 167–171 (1993).

Author Contributions B.J.N., N.T., A.R., D.W.P., Z.J. and J.H. designed and carried out experiments; C.D.M. analysed data; R.S. and T.D.P. supervised research; and B.J.N., N.T., A.R., J.H. and T.D.P. wrote the paper.

Author Information Structural data have been deposited in the Protein Data Bank under accession codes 3DXK (CK-0944636) and 3DXM (CK-0993548). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.D.P. (thomas.pollard@yale.edu).

METHODS

Protein purification. Human, bovine and *S. pombe* Arp2/3 complexes were purified as previously described[11,12,20]. *S. cerevisiae* Arp2/3 complex was purified from strain BN020 (*MAT a or α,ura3-52, his3-200, leu2-3, lys2-801, trp1-1, Δarp2::TRP1, Δarp3::HIS, [pRS315-Arp3], [pRS317-ARP2]*). Cultures were grown overnight at 30 °C to $A_{600} = 1$ and pelleted by centrifuging at 5,000g for 8 min. Cells were washed in lysis buffer (20 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA) and pelleted. The pellet was frozen in liquid nitrogen and stored at −80 °C. Cells were thawed and 1 ml lysis buffer was added per g wet cell pellet. Protease inhibitor tablets (Sigma) were added (one tablet per 100 ml cell suspension) and cells were lysed by five passes at 30,000 p.s.i. on a model 110EH microfluidizer (Microfluidics). The lysate was spun at 30,000g for 20 min and the supernatant was collected and centrifuged for 1 h 15 min at 125,000g. The supernatant was filtered through cheesecloth and proteins were precipitated with 50% saturated aqueous $(NH_4)_2SO_4$. Pelleted proteins were resuspended in 50 ml of PKME (25 mM PIPES pH 7.0, 50 mM KCl, 1 mM EGTA, 3 mM MgCl₂, 1 mM EGTA, 1 mM DTT and 0.1 mM ATP) and dialysed against the same buffer overnight. The dialysate was purified on a GST-N-WASP-VCA affinity column, and Arp2/3 complex-containing fractions were pooled and concentrated to 1 ml before loading onto a Superdex200 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0 and 100 mM NaCl. Peak fractions were pooled and concentrated to approximately 10–20 μM, flash-frozen in liquid nitrogen, and stored at −80 °C. Bovine N-WASP-VCA was constructed by amplifying residues 428–505 from pGEX-2T-NWASP-VCA (a gift of H. Higgs) and cloning into pGV67 using BamHI and EcoRI restriction sites. This vector was used to transform *E. coli* strain BL21(DE3). N-WASP-VCA was expressed and purified as previously described[11]. Chicken skeletal muscle was purified and labelled with pyrene iodoacetamide[21]. Recombinant full length human WASP was purified from 293 cells and recombinant human WASP (residues 105–502) was purified from *E. coli* for the initial screens with HsArp2/3 complex. Recombinant *Listeria* ActA residues 36–170 were expressed as an N-terminal GST fusion and purified from *E. coli*. *S. pombe* formin Cdc12(FH2) domain (residues 973–1390, abbreviated Cdc12(FH2) in text) was purified from *E. coli*.

Actin polymerization assay. We measured the nucleation activity of Arp2/3 complex from the time course of actin polymerization. Polymerization reactions of 100 μl were assembled as previously described[12] and fluorescence measurements were made at 8 s intervals in a 96-well plate using a Gemini XPS spectrofluorimeter (Molecular Devices) with an excitation wavelength of 365 nm and an emission wavelength of 407 nM. The rate of polymerization at each time point was determined by calculating the slope in five-point intervals and multiplying by [total polymer]/ΔRFU, where [total polymer] is the total actin minus the critical concentration (0.1 μM) and ΔRFU is the difference in relative fluorescence units of the reaction upon completion (RFU_{max}) and the initial reaction mix (RFU_{min}). We plotted the maximum elongation rate (R) verses inhibitor concentration ([I]) and fitted the curve using the equation:

$$R = R_{max} - \left[\frac{(R_{max} - R_{min})}{1 + IC_{50}/[I]} \right]$$

Here R_{max} is the maximum rate of elongation with Arp2/3 complex and no inhibitors and R_{min} is the maximum rate of elongation of actin alone. The products of polymerization reactions were diluted into Alexa 488-phalloidin and observed by fluorescence microscopy[16].

Fluorescence anisotropy. The endogenous cysteine (Cys 431) near the end of the amino terminus in purified N-WASP-VCA was labelled with rhodamine maleimide[12]. Fixed concentrations of Rho-N-WASP-VCA were titrated with BtArp2/3 complex and the fluorescence anisotropy measured[12]. Binding constants were determined by fitting the anisotropy curves to the following equation:

$$r = r_f + (r_b - r_f) \left(\frac{(K_d + [R] + [L]) - \sqrt{(K_d + [R] + [L])^2 - 4[R][L]}}{2[L]} \right)$$

where r_f is the signal of the free receptor (R), r_b is the signal of the bound receptor, and [L] is the total concentration of the ligand (species titrated). r_b and K_d were fitted using Kaleidagraph (Synergy Software).

Etheno-ATP fluorescent assay. Binding of etheno-ATP to Arp2/3 complex was performed as described previously with slight modifications[17,18]. Human Arp2/3 complex (0.7 μM) was incubated with 2.5 μM etheno-ATP, 0.2 M acylamide (used as an etheno-ATP fluorescence quencher) in 2 mM Tris-HCl pH 8.0, 50 mM KCl, 0.05 mM EGTA 0.8 mM MgCl₂, 0 and 0.5 mM DTT. Ethano-ATP emission spectra were acquired upon excitation at 340 nm using spectrofluorometer FluoroMax (Horiba JY). Inhibition of Arp2/3 complex by 100 μM CK-636 was performed in the presence or absence of 2.8 μM WASP-VCA.

***Listeria* comet, motility and podosome assays.** SKOV3, THP-1 and *L. monocytogenes* (ATCC 984) cells were from ATCC. SKOV3 cells were cultured in RPMI with 5% FBS without antibiotic at 37 °C and 5% CO₂. *Listeria* were grown in shaking culture in brain heart infusion (Difco Inc., DF0037-15) at 37 °C. SKOV3 cells were infected with *L. monocytogenes*, incubated for 90 min and then treated for 60 min with either compound dissolved in DMSO or a DMSO control before fixing with 2% formaldehyde. Fixed cells were permeabilized with 0.1% Triton-X in PBS at room temperature for 15 min and reacted with a 1:200 dilution of *Listeria* antibody (US Biologics) in PBS for 60 min at room temperature. Antibody was removed and cells were stained with a 1:100 dilution of Alexa Fluor 568 phalloidin and a 1:400 dilution Alexa Fluor 488 goat anti-rabbit secondary antibody (both from Molecular Probes) for 60 min at room temperature. After washing, cells were imaged at 20× with an Axon Instruments automated fluorescence microscope system (Molecular Devices). The THP-1 monocyte cell line was grown in RPMI with 10% FBS/2 mM L-glutamine, differentiated in 50 nM phorbol myristate acetate (Sigma-Aldrich Fluka) for 48 h at 37 °C, trypsinized and plated at 5×10^5 cells ml^{-1} in black Nunc glass-bottom 96-well plates. After incubation at 37 °C for 2 h, cells were treated with compounds for 15 min before fixing and staining with Alexa Fluor 568 phalloidin and DAPI (Sigma-Aldrich Fluka) as above. Supplementary Fig. 2 details the metrics used for podosome quantification. Black molly keratocytes were isolated[19], treated briefly with trypsin and observed by time lapse phase contrast microscopy on plastic Petri dishes at room temperature.

X-ray crystallography. Crystals of BtArp2/3 complex were grown at 4 °C using the hanging drop vapour diffusion method as previously described[7]. Crystals were soaked for 24 h in 18% polyethylene glycol 8000, 50 mM HEPES pH 7.5, 100 mM potassium thiocyanate, 20% glycerol and either 0.5 mM CK-548 or 1 mM CK-636 in DMSO before freezing in liquid nitrogen. Data were collected at beamline X29A at Brookhaven National laboratories and processed with HKL2000[22]. To generate initial models, protein atoms from 1K8K were refined against the data using 30 cycles of minimization in which each subunit was allowed to move as a separate rigid body. CK-548 structure was refined in Refmac[23] with topology files generated in ccp4i sketcher. The CK-636 structure was refined in CNS[8] using topology files generated with the prodrg server[24].

20. Higgs, H. N., Blanchoin, L. & Pollard, T. D. Influence of the C terminus of Wiskott-Aldrich syndrome protein (WASp) and the Arp2/3 complex on actin polymerization. *Biochemistry* **38**, 15212–15222 (1999).
21. Pollard, T. D. Polymerization of ADP-actin. *J. Cell Biol.* **99**, 769–777 (1984).
22. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
23. Dodson, E. J., Winn, M. & Ralph, A. Collaborative Computational Project, number 4: providing programs for protein crystallography. *Methods Enzymol.* **277**, 620–633 (1997).
24. Schuettelkopf, A. W. & van Aalten, D. M. F. PRODRG — a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).

XIAP discriminates between type I and type II FAS-induced apoptosis

Philipp J. Jost[1], Stephanie Grabow[1,2], Daniel Gray[1], Mark D. McKenzie[2,3], Ueli Nachbur[4], David C. S. Huang[1], Philippe Bouillet[1], Helen E. Thomas[3], Christoph Borner[5], John Silke[4], Andreas Strasser[1]* & Thomas Kaufmann[1]*†

FAS (also called APO-1 and CD95) and its physiological ligand, FASL, regulate apoptosis of unwanted or dangerous cells, functioning as a guardian against autoimmunity and cancer development[1–4]. Distinct cell types differ in the mechanisms by which the 'death receptor' FAS triggers their apoptosis[1–4]. In type I cells, such as lymphocytes, activation of 'effector caspases' by FAS-induced activation of caspase-8 suffices for cell killing, whereas in type II cells, including hepatocytes and pancreatic β-cells, caspase cascade amplification through caspase-8-mediated activation of the pro-apoptotic BCL-2 family member BID (BH3 interacting domain death agonist)[5] is essential[6–8]. Here we show that loss of XIAP (X-chromosome linked inhibitor of apoptosis protein)[9,10] function by gene targeting or treatment with a second mitochondria-derived activator of caspases (SMAC[11], also called DIABLO[12]; direct IAP-binding protein with low pI) mimetic drug in mice rendered hepatocytes and β-cells independent of BID for FAS-induced apoptosis. These results show that XIAP is the critical discriminator between type I and type II apoptosis signalling and suggest that IAP inhibitors should be used with caution in cancer patients with underlying liver conditions.

Hepatocytes are sensitive to FASL[13] or agonistic anti-FAS antibodies[14], and observations of invasion of FASL-expressing cytotoxic lymphocytes into hepatic sinusoids implicated FAS-induced apoptosis as a cause of a variety of acute or chronic liver diseases[15]. Caspase-8 (ref. 16) and its activator FADD/MORT1 (FAS-associated death domain protein; ref. 17) are required for FAS-induced apoptosis in all cell types studied so far. Although initial studies produced conflicting results[13,18], it is now clear that amplification of apoptosis signalling through caspase-8-mediated activation of the BH3-only protein BID leading to BAX/BAK-dependent activation of caspase-9 and effector caspases is essential in hepatocytes and pancreatic β-cells (type II cells), but dispensable in lymphocytes (type I cells; Supplementary Fig. 1)[6–8].

It is unclear why FAS activates substantially different apoptotic pathways in different cell types, but it has been postulated that this may be due to differences in the extent of FAS aggregation or internalization, extent of caspase cascade activation, levels of caspase inhibitors (XIAP) and/or abundance of caspase substrates that need to be proteolysed for cells to die[4,18–20]. To explore the differences between type I and type II cells, we compared the levels and activation of apoptosis regulators and the processing of caspase substrates between type I cells, thymocytes, and type II cells, hepatocytes, after FAS stimulation (Supplementary Fig. 2a–c). Processing of caspase-8 into its cleaved form (p18) could be detected as early as 15 min after FAS stimulation by immunoblotting or pull down of the active enzyme with biotinylated X-Val-Ala-Asp-fluoromethylketone (X-VAD-fmk; caspase inhibitor that binds only active caspases) followed

by immunoblotting with a caspase-8-specific antibody (Supplementary Fig. 2a). Caspase-8-mediated BID cleavage and activation of caspase-3 and -7 became evident by ~15 and ~60 min, respectively (Supplementary Fig. 2a, b). Apoptosis induction was comparable in wild-type thymocytes and hepatocytes (Supplementary Fig. 2a–c). As reported previously[7], FASL elicited a similar extent of apoptosis in wild-type and $Bid^{-/-}$ thymocytes in both single-cell suspension and fetal thymic organ culture (FTOC; Supplementary Fig. 2a, c). BID-deficient thymocytes and hepatocytes exhibited normal levels of early caspase-8 activation, but BID-deficient hepatocytes showed considerably less caspase-9 activation and a complete lack of effector caspase activation (Supplementary Fig. 2a, b). The levels of anti-apoptotic BCL-2, BCL-X_L, MCL-1 and the pro-apoptotic SMAC/DIABLO were similar in thymocytes and hepatocytes and remained largely unchanged during FAS activation (Fig. 1a and Supplementary Fig. 2d, e).
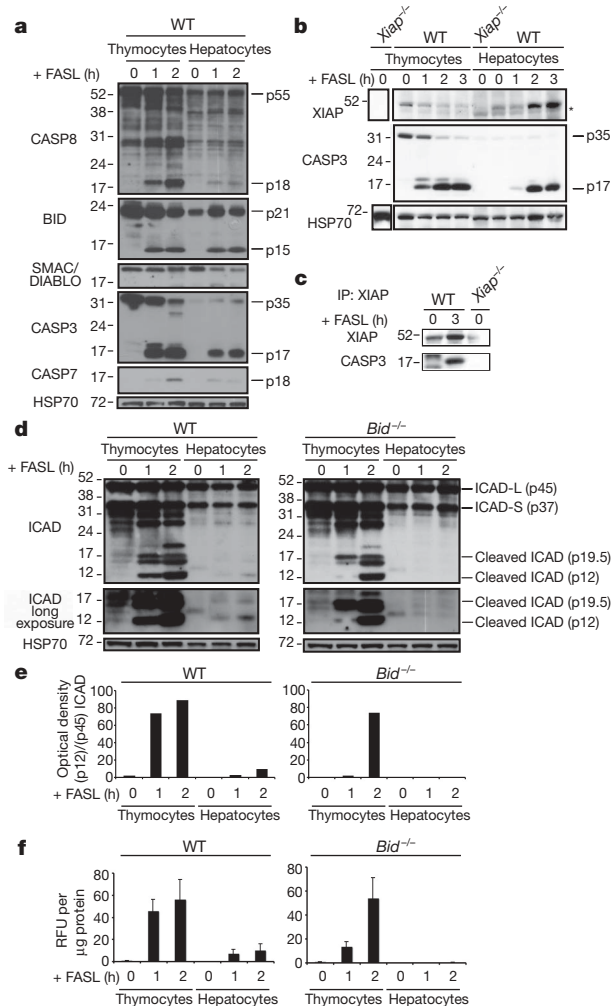
Notably, despite the fact that the basal levels of XIAP were comparable in both cell types, FAS stimulation affected XIAP in substantially different ways in thymocytes compared with hepatocytes. In thymocytes XIAP levels declined rapidly (Fig. 1b), but this could be prevented by inhibition of apoptosis with the caspase blocker qVD-OPh (quinoline-Val-Asp-difluorophenoxymethylketone) (Supplementary Fig. 3a, b), indicating that it is a consequence of cell death. In marked contrast, in hepatocytes treatment with FASL resulted in an increase in XIAP levels, despite efficient initiation of apoptosis (Fig. 1b). This increase in XIAP could be blocked by addition of qVD-OPh or BID deficiency, which both prevent effector caspase activation (Supplementary Fig. 3c, d). Co-immunoprecipitation showed that FAS stimulation caused an increase in the amount of processed caspase-3 associated with XIAP in hepatocytes (Fig. 1c), indicating that XIAP is stabilized by binding and inhibiting activated effector caspases. Furthermore, substantial differences in cleavage of the proto-typical effector caspase substrate ICAD (inhibitor of caspase-activated DNase)[21,22] were detected in thymocytes and hepatocytes. Notably, in wild-type hepatocytes ICAD processing was considerably less efficient than that in wild-type or $Bid^{-/-}$ thymocytes, and in BID-deficient livers no ICAD cleavage could be detected (Fig. 1d, e). Accordingly, fluorogenic (DEVDase) assays showed that FASL-treated wild-type and $Bid^{-/-}$ thymocytes had ~sixfold higher levels of effector caspase activity than wild-type hepatocytes, and no such caspase activity was detected in livers from FASL-injected $Bid^{-/-}$ mice (Fig. 1f). These results indicate that the ratio between activated effector caspases, the caspase inhibitor XIAP and the caspase substrates that need to be proteolysed for a cell to die may determine whether cells undergo type I or type II FAS-induced apoptosis.

Therefore and because XIAP is a potent inhibitor of caspase-9 and effector caspases[9,10], and because XIAP overexpression delayed

Figure 1 | Levels of XIAP, caspase activation and proteolysis of caspase substrates in FASL-treated thymocytes and hepatocytes. a, Lysates from wild-type (WT) thymocytes treated in culture with FASL and livers of wild-type mice injected with FASL were examined by immunoblotting for the expression and processing of caspase-8, BID, SMAC, caspase-3, -7 and HSP70 (loading control). **b,** Lysates from thymocytes and hepatocytes of the indicated genotype treated as described in **a** were examined using antibodies to XIAP, caspase-3 and HSP70 (loading control). Asterisk indicates cross-reactive band. **c,** The binding of active caspase-3 to XIAP in liver lysates described in **b** was examined by immunoprecipitation (IP) with an antibody to XIAP followed by immunoblotting with antibodies to active caspase-3 or XIAP (loading control). **d,** Lysates as described in **a** were examined using antibodies to ICAD and HSP70 (loading control). **e,** Relative optical density of bands in **d** of cleaved ICAD (p12) in relation to total ICAD-L (p45) is shown. **f,** Fluorogenic measurement of effector caspase activity (DEVDase) in lysates as described in **a**. Three independent measurements of two individual mice per genotype and time point are shown. RFU, Relative fluorescence units. Error bars indicate s.e.m.

FAS-induced apoptosis in mouse thymocytes[23], we explored the role of XIAP in FAS-induced hepatocyte killing *in vivo*. As reported previously[13,14], wild-type mice succumbed to injection with FASL or anti-FAS antibody within 2.5–3 (Fig. 2a) or 3–4 h, respectively. Interestingly, these treatments killed XIAP-deficient mice considerably earlier (1–2 h for FASL, Fig. 2a; ~2 h for anti-FAS antibody). Sick animals of both genotypes presented with abnormally elevated serum levels of the liver-associated transaminases ALT (alanine aminotransferase) and AST (aspartate aminotransferase; Fig. 2b) and severe liver damage (Fig. 2g, h). In contrast, loss of XIAP did not increase FAS-mediated apoptosis in thymocytes (type I cells; data not shown and ref. 24). Thus, XIAP is a critical attenuator of FAS-induced hepatocyte apoptosis.
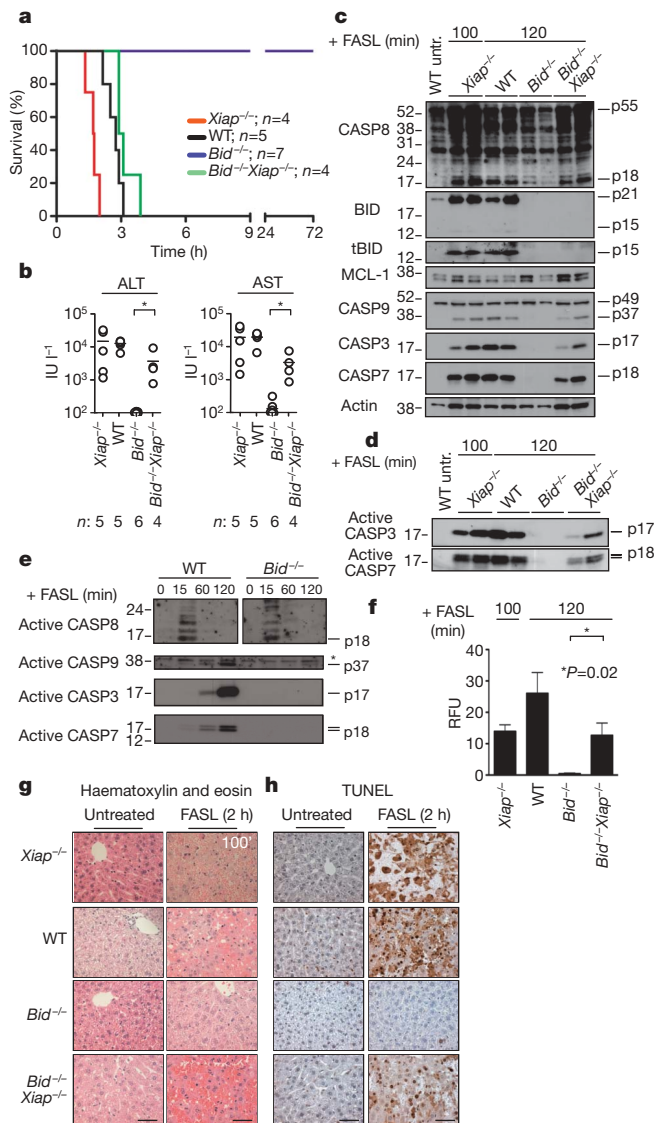
Figure 2 | Loss of XIAP re-sensitises $Bid^{-/-}$ mice to FASL-induced fatal hepatitis. a, Survival of wild-type, $Xiap^{-/-}$, $Bid^{-/-}$ and $Bid^{-/-}Xiap^{-/-}$ mice injected intravenously with FASL is shown. Wild type versus $Bid^{-/-}$, $P < 0.0003$; $Bid^{-/-}$ versus $Bid^{-/-}Xiap^{-/-}$, $P = 0.0005$; wild type versus $Bid^{-/-}Xiap^{-/-}$, $P = 0.1133$. **b,** Serum levels of ALT and AST in mice (**a**) after 120 min of FASL treatment (100 min for $Xiap^{-/-}$ mice). Horizontal bars indicate the mean; n, numbers of mice analysed. Asterisk, $P = 0.042$ (ALT) or 0.022 (AST). **c,** Levels and processing of caspase-3, -7, -8, -9, BID (tBID exposed longer), MCL-1 and β-actin (loading control) in livers of FASL-injected mice (**a**) was examined by immunoblotting. **d,** Activation of caspase-3 and -7 in livers of FASL-injected mice (**a**) was examined by pull down with X-VAD-fmk followed by immunoblotting. **e,** Activation of caspase-3, -7, -8 and -9 in liver extracts from FASL-injected wild-type or $Bid^{-/-}$ mice (**a**) was examined as in **d**. Asterisk indicates a cross-reactive band. **f,** Fluorogenic measurement of effector caspase activity (DEVDase) in the livers of FASL-injected mice (**a**). Data represent means and s.e.m. from four independent measurements of two mice per genotype and time point. RFU, Relative fluorescence units; asterisk, $P = 0.02$. **g, h,** Histology (**g**) and TUNEL staining (**h**) of liver sections from FASL-injected mice (**a**) killed as in **b** and controls. Images are representative of at least three mice per genotype and time point. Scale bars, 50 μm.

To examine whether XIAP is the crucial discriminator between type I and type II FAS-induced apoptosis signalling, we generated BID/XIAP double-deficient mice, which exhibited no obvious abnormalities under unstressed conditions. As previously reported[6,7], BID-deficient mice survived injection with FASL or anti-FAS antibody without developing signs of illness and showed no increase in serum ALT or AST levels or disruption of liver architecture (Fig. 2 and

Supplementary Fig. 4a). In contrast, all BID/XIAP double-deficient animals succumbed to these treatments (albeit slightly later than wild-type mice, ~3–3.5 versus 2.5–3 h), presenting with abnormally elevated ALT/AST serum levels, severe liver damage and classical signs of hepatocyte apoptosis, including chromatin condensation and DNA fragmentation (Fig. 2a, b, g, h and Supplementary Fig. 4a). Immunoblotting demonstrated that FAS stimulation caused similar levels of caspase-8 processing in livers of mice of all genotypes, but caspase-9 and effector caspases were only activated in livers of the sensitive animals (wild type, $Xiap^{-/-}$ and $Bid^{-/-}Xiap^{-/-}$), but not in those from the resistant $Bid^{-/-}$ mice (Fig. 2c). This was confirmed by pull down of active caspase-8, -9, -3 and -7 with biotinylated X-VAD-fmk followed by immunoblotting with caspase-specific antibodies (Fig. 2d, e) and by fluorogenic effector caspase activity (DEVDase) assays (Fig. 2f).

In accordance with our analyses of hepatocytes, BID deficiency protected pancreatic β-cells, which also undergo type II FAS-induced apoptosis[8], from FASL, but concomitant loss of XIAP (in $Bid^{-/-}Xiap^{-/-}$ mice) restored sensitivity (Supplementary Fig. 5).

Remarkably, pre-treatment with the caspase inhibitor qVD-OPh significantly ($P = 0.0042$) protected $Bid^{-/-}Xiap^{-/-}$ mice from FASL-induced fatal hepatitis and these animals contained lower ALT ($P = 0.078$) and AST ($P = 0.024$) levels in their sera and suffered less hepatocyte destruction compared to $Bid^{-/-}Xiap^{-/-}$ animals that had only been injected with FASL (Fig. 3a–d and Supplementary Fig. 4b). As expected, pre-treatment with qVD-OPh inhibited FASL-induced processing of caspase-8 and subsequent activation of BID and caspase-3, -7 and -9 (Fig. 3e). Thus, FAS stimulation kills BID/XIAP double-deficient hepatocytes through caspase-dependent apoptosis and not some caspase-independent process. Collectively, these results show that XIAP constitutes the critical regulator that discriminates between type I and type II FAS-induced apoptosis signalling.

An important question is whether the activation of caspase-9 by cytochrome c and APAF1 (apoptotic protease-activating factor 1) is necessary for type II FAS-induced apoptosis, or whether mitochondrial release of XIAP antagonists (for example, SMAC), leading to liberation of active effector caspases, is sufficient. Notably, 120 min after FASL injection, apoptotic nuclei (Fig. 2h) and caspase-3 and -7 activation (Fig. 2c, d) were readily detectable in livers from $Bid^{-/-}Xiap^{-/-}$ mice, although their levels of cleaved caspase-9 were considerably lower than that of wild-type hepatocytes (Fig. 2c). The caspase-9 processing seen in $Bid^{-/-}Xiap^{-/-}$ livers (Fig. 2c) might thus be a consequence of pro-caspase-9 cleavage by already activated effector caspases, indicating that caspase-9 may not be necessary for effector caspase activation in FAS-induced hepatocyte apoptosis. Consistent with this hypothesis, mice harbouring a mutation in cytochrome c that prevents apoptosome formation (and hence caspase-9 activation) showed normal sensitivity to FAS-induced hepatitis[25], whereas combined loss of caspases-3 and -7 rendered cells resistant to FASL[26].

XIAP and related IAPs (cIAP1/2) are inhibited by SMAC[11,12] and HTRA2 (also called OMI) proteins that are released from the mitochondria during apoptosis[9,10]. SMAC mimetics are currently in development for cancer therapy and were shown to kill tumour cells (at least in part) by enhancing autocrine TNF-α–TNFR1 signalling[27–30]. We examined the impact of the IAP inhibitor BV6 (ref. 27) on FASL-induced hepatocyte killing. Wild-type mice injected with BV6 plus FASL died significantly earlier than wild-type mice treated with FASL alone (Supplementary Fig. 6), indeed at a time comparable to FASL-injected $Xiap^{-/-}$ mice. Remarkably, all $Bid^{-/-}$ mice injected with FASL plus BV6 rapidly developed fatal hepatitis (Fig. 4a), presenting with abnormally increased serum ALT and AST levels (Fig. 4b) and extensive hepatocyte destruction (Fig. 4c, d and Supplementary Fig. 4c), whereas all $Bid^{-/-}$ mice injected with FASL alone or BV6 alone remained healthy. Despite initial activation of caspase-8 in all FASL-treated animals, significant processing of effector caspases was only seen in $Bid^{-/-}$ mice that were injected with FASL plus BV6, but not in those treated with FASL or BV6 alone
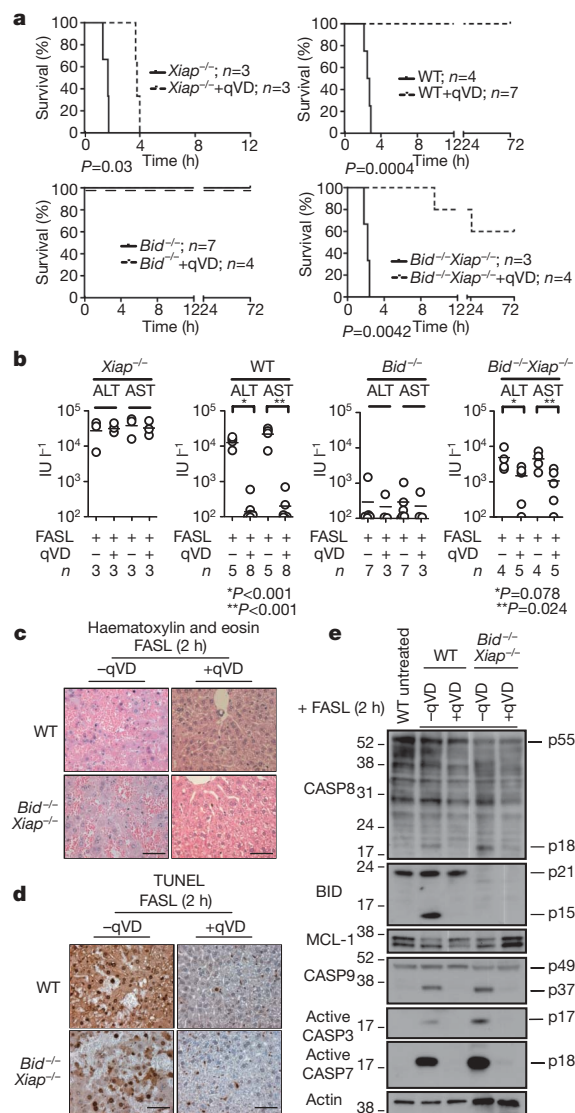
Figure 3 | Caspase inhibitors protect $Bid^{-/-}Xiap^{-/-}$ hepatocytes from FASL-induced apoptosis. a, Survival of FASL-injected (Fig. 2a) wild-type, $Xiap^{-/-}$, $Bid^{-/-}$ and $Bid^{-/-}Xiap^{-/-}$ mice with or without pre-treatment with the caspase inhibitor qVD-OPh (qVD). **b,** Serum levels of ALT and AST in mice (**a**) killed after 120 min (100 min for $Xiap^{-/-}$). Horizontal bars indicate the mean; n, numbers of mice analysed. **c, d,** Histology (**c**) and TUNEL staining (**d**) of liver sections from mice (**a**) killed as in **b**. Images are representative of at least three mice per genotype and time point. Scale bars, 50 μm. **e,** Levels and processing of caspase-3, -7, -8, -9, BID, MCL-1 and β-actin (loading control) in extracts of livers from mice (**a**) was examined by immunoblotting.

(Fig. 4e). Pre-treatment with qVD-OPh protected $Bid^{-/-}$ mice from FASL-plus-BV6-induced fatal hepatitis (Fig. 4a) and this was reflected by lower ALT/AST serum levels compared to those seen in animals treated with FASL alone (Fig. 4b). Histological examination and TdT-mediated dUTP nick end labelling (TUNEL) staining revealed normal liver architecture and only few apoptotic hepatocytes in mice treated with FASL, BV6 and qVD-OPh combined (Fig. 4c, d and Supplementary Fig. 4c). This demonstrates that the combination of FASL plus BV6 triggers caspase-dependent hepatocyte apoptosis.

BV6 treatment could theoretically promote FASL-induced hepatocyte killing by blocking cIAP1 and/or cIAP2, which would be expected to require autocrine/paracrine TNF-α–TNFR1 signalling[27–30]. Treatment with FASL, BV6 or both had, however, no impact on the levels of TNF-α in livers (Supplementary Figs 6c and 7c) and, pertinently, TNF-α-neutralizing antibodies had no effect on FASL-plus-BV6-induced
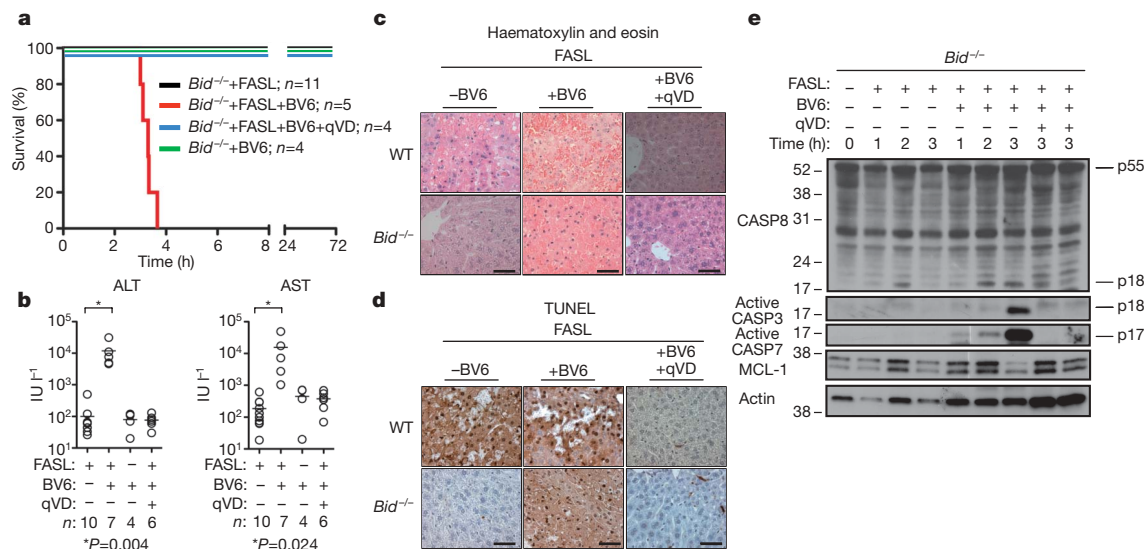
Figure 4 | The SMAC-mimetic drug BV6 sensitises BID-deficient mice to FASL-induced hepatocyte destruction. a, Survival of FASL-injected (Fig. 2a) $Bid^{-/-}$ mice with or without pre-treatment with BV6 and/or qVD-OPh as indicated is shown. **b,** Serum levels of ALT and AST after 120 min. Horizontal bars indicate the mean; n, numbers of mice analysed. Asterisk, $P = 0.004$ (ALT) or 0.024 (AST). **c, d,** Histology (**c**) and TUNEL staining (**d**) of liver sections from mice (**a**) killed after 180 min (120 min for animals treated with BV6 alone). Images are representative of at least three mice per genotype and time point. Scale bars, 50 μm. **e,** Levels and processing of caspase-3, -7, -8, MCL-1 and β-actin (loading control) in liver extracts of $Bid^{-/-}$ mice (**a**) killed at the indicated time points were examined by immunoblotting.

hepatocyte destruction in $Bid^{-/-}$ mice (Supplementary Fig. 7a, b) although they could prevent TNF-α-induced killing of L929 cells and lipopolysaccharide (LPS)-plus-galactosamine (GalN)-induced fatal hepatitis, a TNF-α–TNFR1-dependent process[4] (Supplementary Fig. 8). Thus, BV6 does not need to activate TNF–TNFR1 signalling to promote FASL-induced hepatocyte killing, but functions by blocking the anti-apoptotic activity of XIAP. The notion that XIAP rather than cIAP1/cIAP2 is the critical target of BV6 in FAS-induced hepatocyte apoptosis is underscored by the finding that only $Xiap^{-/-}$ mice (Fig. 2a and Supplementary Fig. 6a), but not mice lacking cIAP1, cIAP2 or both in their livers, succumbed more rapidly to FASL-induced hepatitis than wild-type animals (not shown). The importance of XIAP in regulating FAS-induced apoptosis predicts that the endogenous IAP antagonists SMAC and/or HTRA2 have critical roles in this process. Because $Bid^{-/-} Xiap^{-/-}$ mice succumb to FASL-induced hepatitis more slowly than wild-type animals, apoptosis inhibitors in addition to XIAP, such as ML-IAP[10], may contribute to the regulation of this process. Finally, although IAP inhibition represents a promising strategy for cancer therapy, our data indicate that IAP inhibitors should be used with caution in combination with liver toxic chemotherapy or patients with underlying liver conditions.

METHODS SUMMARY

FASL-induced hepatitis. Mice were injected intravenously with 0.25 mg per kg body weight recombinant soluble FASL (Flag-tagged FASL, Alexis Biochemicals) crosslinked with 2 μg anti-Flag antibody (M2, Sigma-Aldrich) per μg FASL. The pan-caspase inhibitor qVD-OPh (MP Biomedicals) was injected at a concentration of 20 mg per kg body weight intraperitoneally 30 min before FASL administration. The SMAC mimetic BV6 (a gift from Genentech) was injected intraperitoneally (10 mg per kg body weight) 30 min before FASL administration. Mice were killed at various time points, bled for serum analysis of the liver-associated transaminases ALT and AST and the livers were removed for biochemical and histological analyses. Statistical analyses were performed applying a two-tailed unpaired Student's t-test. Statistical analysis of animal survival was performed using a Log-rank (Mantel–Cox) test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Nagata, S. FAS ligand-induced apoptosis. *Annu. Rev. Genet.* **33**, 29–55 (1999).

2. Krammer, P. H. CD95's deadly mission in the immune system. *Nature* **407**, 789–795 (2000).
3. Peter, M. E. *et al.* The CD95 receptor: apoptosis revisited. *Cell* **129**, 447–450 (2007).
4. Strasser, A., Jost, P. J. & Nagata, S. The many roles of FAS receptor signaling in the immune system. *Immunity* **30**, 180–192 (2009).
5. Wang, K., Yin, X.-M., Chao, D. T., Milliman, C. L. & Korsmeyer, S. J. BID: a novel BH3 domain-only death agonist. *Genes Dev.* **10**, 2859–2869 (1996).
6. Yin, X.-M. *et al.* Bid-deficient mice are resistant to Fas-induced hepatocellular apoptosis. *Nature* **400**, 886–891 (1999).
7. Kaufmann, T. *et al.* The BH3-only protein Bid is dispensable for DNA damage- and replicative stress-induced apoptosis or cell-cycle arrest. *Cell* **129**, 423–433 (2007).
8. McKenzie, M. D. *et al.* Proapoptotic BH3-only protein Bid is essential for death receptor-induced apoptosis of pancreatic β-cells. *Diabetes* **57**, 1284–1292 (2008).
9. Holcik, M. & Korneluk, R. G. XIAP, the guardian angel. *Nature Rev. Mol. Cell Biol.* **2**, 550–556 (2001).
10. Vaux, D. L. & Silke, J. IAPs, RINGs and ubiquitylation. *Nature Rev. Mol. Cell Biol.* **6**, 287–297 (2005).
11. Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
12. Verhagen, A. M. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing inhibitor of apoptosis (IAP) proteins. *Cell* **102**, 43–53 (2000).
13. Huang, D. C. *et al.* Activation of Fas by FasL induces apoptosis by a mechanism that cannot be blocked by Bcl-2 or Bcl-x$_L$. *Proc. Natl Acad. Sci. USA* **96**, 14871–14876 (1999).
14. Ogasawara, J. *et al.* Lethal effect of the anti-Fas antibody in mice. *Nature* **364**, 806–809 (1993).
15. Malhi, H. & Gores, G. J. Cellular and molecular mechanisms of liver injury. *Gastroenterology* **134**, 1641–1654 (2008).
16. Varfolomeev, E. E. *et al.* Targeted disruption of the mouse *Caspase 8* gene ablates cell death induction by the TNF receptors, Fas/Apo1, and DR3 and is lethal prenatally. *Immunity* **9**, 267–276 (1998).
17. Newton, K., Harris, A. W., Bath, M. L., Smith, K. G. C. & Strasser, A. A dominant interfering mutant of FADD/Mort1 enhances deletion of autoreactive thymocytes and inhibits proliferation of mature T lymphocytes. *EMBO J.* **17**, 706–718 (1998).
18. Scaffidi, C. *et al.* Two CD95 (APO-1/Fas) signaling pathways. *EMBO J.* **17**, 1675–1687 (1998).
19. Algeciras-Schimnich, A. *et al.* Molecular ordering of the initial signaling events of CD95. *Mol. Cell. Biol.* **22**, 207–220 (2002).
20. Li, S. *et al.* Relief of extrinsic pathway inhibition by the Bid-dependent mitochondrial release of Smac in Fas-mediated hepatocyte apoptosis. *J. Biol. Chem.* **277**, 26912–26920 (2002).
21. Liu, X., Zou, H., Slaughter, C. & Wang, X. DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* **89**, 175–184 (1997).
22. Enari, M. *et al.* A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43–50 (1998).

23. Conte, D., Liston, P., Wong, J. W., Wright, K. E. & Korneluk, R. G. Thymocyte-targeted overexpression of *xiap* transgene disrupts T lymphoid apoptosis and maturation. *Proc. Natl Acad. Sci. USA* **98**, 5049–5054 (2001).

24. Harlin, H., Reffey, S. B., Duckett, C. S., Lindsten, T. & Thompson, C. B. Characterization of XIAP-deficient mice. *Mol. Cell. Biol.* **21**, 3604–3608 (2001).

25. Hao, Z. *et al.* Specific ablation of the apoptotic functions of cytochrome *c* reveals a differential requirement for cytochrome *c* and Apaf-1 in apoptosis. *Cell* **121**, 579–591 (2005).

26. Lakhani, S. A. *et al.* Caspases 3 and 7: key mediators of mitochondrial events of apoptosis. *Science* **311**, 847–851 (2006).

27. Varfolomeev, E. *et al.* IAP antagonists induce autoubiquitination of c-IAPs, NF-κB activation, and TNFα-dependent apoptosis. *Cell* **131**, 669–681 (2007).

28. Vince, J. E. *et al.* IAP antagonists target cIAP1 to induce TNFα-dependent apoptosis. *Cell* **131**, 682–693 (2007).

29. Petersen, S. L. *et al.* Autocrine TNFα signaling renders human cancer cells susceptible to Smac-mimetic-induced apoptosis. *Cancer Cell* **12**, 445–456 (2007).

30. Gaither, A. *et al.* A Smac mimetic rescue screen reveals roles for inhibitor of apoptosis proteins in tumor necrosis factor-α signaling. *Cancer Res.* **67**, 11493–11498 (2007).

METHODS

Mice. The generation and genotyping of $Bid^{-/-}$ mice[31], generated on an inbred C57BL/6 background using C57BL/6-derived embryonic stem (ES) cells, and $Xiap^{-/-}$ mice[32], generated on a mixed C57BL/6x129SV background using 129SV-derived embryonic stem cells and backcrossed onto C57BL/6 for more than ten generations, have been described previously[31,32]. The $Xiap$ gene is located on the X-chromosome; however, both female ($Xiap^{-/-}$) and male ($Xiap^{y/-}$) mice used for experiments are denoted as $Xiap^{-/-}$ in the text and figures. Inter-crossing $Bid^{-/-}$ with $Xiap^{-/-}$ mice generated BID/XIAP double-deficient mice. All experiments with mice were performed according to the guidelines of the animal ethics committees of the Melbourne Health Research Directorate.

Anti-FAS-antibody-induced hepatitis. Mice were injected intravenously with 0.25 mg per kg body weight of the agonistic hamster monoclonal anti-mouse FAS antibody (clone Jo2, no azide plus low endotoxin formulation, BD Pharmingen).

FASL-induced apoptosis assay of thymocytes *ex vivo*. Single cell suspensions of thymocytes were prepared from freshly excised thymi and cultured in DMEM supplemented with 100 U ml^{-1} penicillin, 100 µg ml^{-1} streptomycin, 2 mM glutamine, 10% fetal calf serum (FCS, Bovogen Biologicals) at 37 °C in 5% CO$_2$. Thymocytes were treated with 100 ng ml^{-1} recombinant soluble FASL (Flag-tagged FASL, Alexis Biochemicals) that had been crosslinked with 2 µg ml^{-1} anti-Flag antibody (M2, Sigma-Aldrich).

Lipopolysaccharide-plus-galactosamine-induced hepatitis. Mice were injected intraperitoneally with 100 ng lipopolysaccharide (LPS, DIFCO) in the presence of 20 mg of the liver-specific transcriptional inhibitor D-(+)-galactosamine (GalN, Sigma-Aldrich) per 20 g body weight of mouse.

Blockade of TNF-α-induced fatal hepatitis by TNF-α-neutralizing antibody *in vivo*. Mice were injected intraperitoneally with 250 µg TNF-α-neutralizing antibody XT-22 or an immunoglobulin isotype-matched control monoclonal antibody, GL113, per 20 g body weight of mouse 30 min before injection of lipopolysaccharide plus galactosamine or FASL.

Immunoblotting. Liver lysates were prepared in ONYX buffer containing 20 mM Tris/HCl pH 7.4, 135 mM NaCl, 1.5 mM MgCl$_2$, 1 mM EGTA, 10% glycerol, 1% Triton X-100, 500 µg ml^{-1} Pefabloc (AEBSF), 1 µg ml^{-1} each of leupeptin, aprotinin and pepstatin, 100 µg ml^{-1} soybean trypsin inhibitor, 2 µg ml^{-1} E64 or alternatively in RIPA buffer containing 150 mM NaCl, 1% Triton X-100, 0.5% deoxycholic acid, 0.1% SDS, 50 mM Tris/HCl pH 8.0 supplemented with protease inhibitors as indicated above. Proteins in liver lysates were size-separated on pre-cast 12% or 4–20% SDS–PAGE gradient gels (Invitrogen). Membranes were probed with the monoclonal antibodies rat anti-BID (clone 2D1 (ref. 31), rat anti-caspase-8 (clone 1G12, Alexis Biochemicals[33]), mouse anti-XIAP (clone 2F1, Abcam), rat anti-MCL-1 (clone 19C4-15), mouse anti-β-actin (used as a loading control; clone AC40, Sigma-Aldrich), hamster anti-mouse BCL-2 (clone 3F11), mouse anti-HSP70 (N6, gift from R. Anderson) or polyclonal rabbit antibodies specific for caspase-3 (Cell Signaling), active caspase-7 (p18, Cell Signaling), caspase-9 (gift from Y. Lazebnik), N-terminus BAK (Millipore), N-terminus BAX (Millipore), SMAC (clone 7, BD Transduction Laboratories) or anti-ICAD (gift from S. Nagata). Optical densitometry was performed on a GS-800 Calibrated Densitometer (Bio-Rad) using Quantity One 4.6.1 software (Bio-Rad).

Pull down of active caspases and co-immunoprecipitation plus immunoblot analysis. Active caspases from liver lysates were irreversibly bound to the pan-caspase inhibitor biotin-X-VAD-fmk (Calbiochem) added at a final concentration of 2.5 µM for 30 min at 37 °C. Alternatively, in liver lysates of mice the binding of active caspase-3 to XIAP was tested by co-immunoprecipitation with an antibody to XIAP (clone 2F1, 2 µg ml^{-1}) and immunoblotting with antibodies to active caspase-3. Streptavidin Sepharose or Protein G Sepharose (GE Healthcare) was added as a 50% slurry in ONYX buffer and the mixture incubated for at least 2 h on a rotating wheel at 4 °C. The Sepharose beads were pelleted by centrifugation and washed three times in ONYX buffer before boiling for 5 min in reducing SDS–PAGE sample buffer. Proteins in boiled samples were separated on an SDS–PAGE gel and probed for various caspases by immunoblotting using the caspase-specific antibodies listed above.

Fluorogenic caspase activity assay. Cell lysates were prepared in RIPA buffer. Cleavage assays were set up with 8 µl lysate plus 50 µM DEVD-AMC (Asp-Glu-Val-Asp-7-amino-4-methylcoumarin; Bachem) in a total volume of 100 µl caspase cleavage assay buffer (0.1 M HEPES pH 7.5, 10% sucrose, 0.1% CHAPS, 10 mM dithiothreitol) and fluorescent emission measured at 37 °C in a Spectra Max2 (Molecular Devices) fluorometer at 37 °C for 30 min with individual readings made every 20 s (excitation filter 360 nm; emission filter 465 nm). Data are expressed as relative rates of DEVD-AMC cleavage from four independent experiments.

Histology and TUNEL staining. Livers of treated mice were excised at indicated time points and fixed in 10% buffered formalin solution before preparation of histological sections and staining with haematoxylin and eosin. Tissue sections for TUNEL analysis were de-paraffinized before proteinase K treatment (20 µg ml^{-1}). Endogenous peroxidases were blocked by incubating the sections in 3% hydrogen peroxide in methanol for 5 min at room temperature. Sections were washed once with PBS and once with distilled water and subsequently incubated in TdT buffer for 10 min at room temperature, followed by washing twice with PBS. Tissue sections were then incubated in 0.6 U µl^{-1} terminal deoxynucleotidyl transferase (TdT, Promega), 20 µM biotin-16-dUTP (Roche), 1 mM CoCl$_2$ (Sigma-Aldrich) in terminal transferase buffer (Promega) for 1 h at 37 °C. After three washes, sections were blocked with 2% FCS (Bovogen Biologicals) in PBS for 10 min at room temperature. Blocking solution was removed and Vectastain ABC (Vector Laboratories) reagents applied according to the manufacturer's instructions. Sections were then washed three times with PBS and once with distilled water followed by 3,3′-diaminobenzidine (DAB) staining according to the manufacturer's instructions. Sections were finally washed twice with distilled water and counterstained with haematoxylin.

TNF-α detection by enzyme-linked immunosorbent assay (ELISA). Measurement of TNF-α levels was performed in a 96-well round-bottom plate coated overnight with 100 µl of TNF-α capture antibody solution at 0.8 µg ml^{-1} in PBS followed by three washes with 0.05% PBS/Tween 20. The wells were blocked with 1% BSA in PBS for 1 h at room temperature and three washes with PBS/Tween 20. Samples (100 µl) or standard (in 1% BSA/PBS) were added. Plates were incubated for 2 h at room temperature, followed by three washes with PBS/Tween 20. Detection antibodies (100 µl biotinylated goat anti-mouse TNF-α antibodies at 200 ng ml^{-1} in 1% BSA/PBS) were added to each well and plates incubated for 2 h at room temperature. After three washes with PBS/Tween 20, 100 µl streptavidin-horseradish peroxidase (HRP) solution (diluted 1:200 in 1% BSA/PBS) was added and the plates incubated for 20 min at room temperature. After three washes with PBS/Tween 20, substrate solution and 20 min later stop solution were added according to the manufacturer's instructions (R&D Systems, catalogue numbers DY999 and DY994). The colorimetric reaction was measured at 450 nm on a Multiscan Ascent ELISA plate reader (Thermo Labsystems).

Fetal thymic organ culture. Thymic lobes were dissected from wild-type or $Bid^{-/-}$ embryos at day 15 of gestation (day of vaginal plug formation designated as day 0) and cultured as described previously[34] in RPMI medium supplemented with 15 mM HEPES and 10% (v/v) FCS at 37 °C in 5% CO$_2$. After 5 days of culture, lobes were transferred into 'hanging drop' cultures in Terasaki plates (NUNC) with medium containing 1 µg ml^{-1} FASL cross-linked with 20 µg ml^{-1} anti-Flag antibody and cultured for a further 4, 8 or 16 h. Single-cell suspensions of pools of 10–20 lobes were counted and sampled for preparation of cell lysates for immunoblotting and flow cytometric analysis of cell survival, using staining with propidium iodide plus fluorochrome-conjugated antibodies to CD4 and CD8.

Pancreatic islet isolation and cell death assay. Pancreatic islets were isolated from mice using collagenase P (Roche) and histopaque-1077 density gradients (Sigma-Aldrich) as described previously[35]. Islets were washed, handpicked and cultured overnight at 37 °C and 5% CO$_2$ in CMRL medium-1066 (Gibco products; Invitrogen) supplemented with 100 U ml^{-1} penicillin, 100 µg ml^{-1} streptomycin, 2 mM glutamine and 10% FCS (JRH Biosciences). The day after isolation, 100 uniformly sized islets (excluding very large or necrotic islets) per sample were handpicked into 3.5-cm Petri dishes containing 1.1 ml complete CMRL medium and cultured for 6 days with and without 100 ng ml^{-1} Flag-tagged FASL crosslinked with anti-Flag antibody (2 µg ml^{-1}) and cytokines (1 U ml^{-1} recombinant murine IFN-γ and 150 U ml^{-1} recombinant murine IL-1β (R&D systems)). At the end of the culture period, non-attached cells and islets were transferred into polypropylene tubes and washed in PBS and dispersed (0.1 mg ml^{-1} bovine trypsin (Calbiochem) and 2 mM EDTA in PBS) for 5 min at 37 °C. Cells were then washed in PBS and resuspended in 250 µl hypotonic buffer containing 50 µg ml^{-1} propidium iodide (Sigma-Aldrich), 0.1% sodium citrate and 0.1% Triton X-100, which stains nuclear DNA. Cells were then analysed on a FACSCalibur (Becton Dickinson). Apoptotic cells were identified by their sub-diploid (<2C) DNA content as previously described[36].

***In vitro* cell culture.** Mouse fibroblastoid L929 cells were maintained in DMEM supplemented with 10% FCS, 100 U ml^{-1} penicillin, 100 µg ml^{-1} streptomycin, 2 mM glutamine and used at 80% confluency for cell death measurement after incubation with soluble, recombinant murine MegaTNF-α (Alexis Biochemicals) at a concentration of 100 ng ml^{-1}. The TNF-α-neutralizing antibody XT-22 and an immunoglobulin isotype-control monoclonal antibody, GL113, were used at a concentration of 1 µg ml^{-1} and added to L929 cells 30 min before addition of MegaTNF-α. Cell death was measured by FACS analysis in a FACSCalibur (Becton Dickinson) after staining with propidium iodide.

31. Kaufmann, T. *et al.* The BH3-only protein Bid is dispensable for DNA damage- and replicative stress-induced apoptosis or cell-cycle arrest. *Cell* **129**, 423–433 (2007).

32. Olayioye, M. A. *et al.* XIAP-deficiency leads to delayed lobuloalveolar development in the mammary gland. *Cell Death Differ.* **12**, 87–90 (2005).

33. O'Reilly, L. A. *et al.* Modifications and intracellular trafficking of FADD/MORT1 and caspase-8 after stimulation of T lymphocytes. *Cell Death Differ.* **11**, 724–736 (2004).

34. Waanders, G. A. & Boyd, R. L. The effects of interleukin 2 on early and late thymocyte differentiation in foetal thymus organ culture. *Int. Immunol.* **2**, 461–468 (1990).

35. McKenzie, M. D. *et al.* Perforin and Fas induced by IFN and TNF mediate beta cell death by OT-I CTL. *Int. Immunol.* **18**, 837–846 (2006).

36. Nicoletti, I., Migliorati, G., Pagliacci, M. C., Grignani, F. & Riccardi, C. A rapid and simple method for measuring thymocyte apoptosis by propidium iodide staining and flow cytometry. *J. Immunol. Methods* **139**, 271–279 (1991).

LETTERS

Structure of a prokaryotic virtual proton pump at 3.2 Å resolution

Yiling Fang[1], Hariharan Jayaram[1], Tania Shane[1], Ludmila Kolmakova-Partensky[1], Fang Wu[1], Carole Williams[1], Yong Xiong[2] & Christopher Miller[1]

To reach the mammalian gut, enteric bacteria must pass through the stomach. Many such organisms survive exposure to the harsh gastric environment (pH 1.5–4) by mounting extreme acid-resistance responses, one of which, the arginine-dependent system of *Escherichia coli*, has been studied at levels of cellular physiology, molecular genetics and protein biochemistry[1–7]. This multiprotein system keeps the cytoplasm above pH 5 during acid challenge by continually pumping protons out of the cell using the free energy of arginine decarboxylation. At the heart of the process is a 'virtual proton pump'[8] in the inner membrane, called AdiC[3,4], that imports L-arginine from the gastric juice and exports its decarboxylation product agmatine. AdiC belongs to the APC superfamily of membrane proteins[6,7,9], which transports amino acids, polyamines and organic cations in a multitude of biological roles, including delivery of arginine for nitric oxide synthesis[10], facilitation of insulin release from pancreatic β-cells[11], and, when inappropriately overexpressed, provisioning of certain fast-growing neoplastic cells with amino acids[12,13]. High-resolution structures and detailed transport mechanisms of APC transporters are currently unknown. Here we describe a crystal structure of AdiC at 3.2 Å resolution. The protein is captured in an outward-open, substrate-free conformation with transmembrane architecture remarkably similar to that seen in four other families of apparently unrelated transport proteins.

The proton-extrusion function of AdiC, formally an arginine/agmatine antiporter, arises from linkage of transport to substrate decarboxylation (Fig. 1a), a reaction carried out by a separate enzyme, AdiA, wherein an aqueous proton replaces the α-carboxyl group of L-arginine (Arg) to form a C–H bond on agmatine (Agm); export of this 'virtual proton' counteracts cytoplasmic acidification that would otherwise occur at low extracellular pH. Physiological imperatives demand that AdiC specifically imports the deprotonated-carboxylate Arg$^+$ form, which below pH 2 represents a minor fraction of extracellular Arg[2,6]; transport of the predominant Arg^{2+} form would produce a futile cycle useless for acid resistance, as a carboxyl proton would enter for each virtual proton pumped out. Transport of the Arg^{2+} form cannot be directly measured in our *E. coli* AdiC-reconstituted liposomes at pH 6, but argininamide^{2+} (Arg-NH$_2$), an isosteric proxy for protonated Arg^{2+}, is readily tested. In these experiments (Fig. 1b), [^{14}C]Arg is allowed to accumulate into liposomes preloaded with Arg. Radiolabel is then expelled from the liposomes by addition of a low concentration (0.1 mM) of unlabelled Arg to the external medium, according to competition for uptake between labelled and unlabelled substrate. A repeat of this assay with Arg-NH$_2$ shows that this analogue is, as expected, a poor substrate for transport, because >30-fold higher concentration is required to mimic competition by Arg. A negatively charged carboxylate cannot be a general requirement for transport, however, as the physiological

substrate Agm is an excellent substrate[6]. Figure 1b also presents negative transport controls using mutations at certain conserved aromatic residues that inhibit activity without affecting protein assembly[7]. In contrast to many broad-spectrum amino acid transporters of the APC superfamily, AdiC and homologous virtual proton pumps show
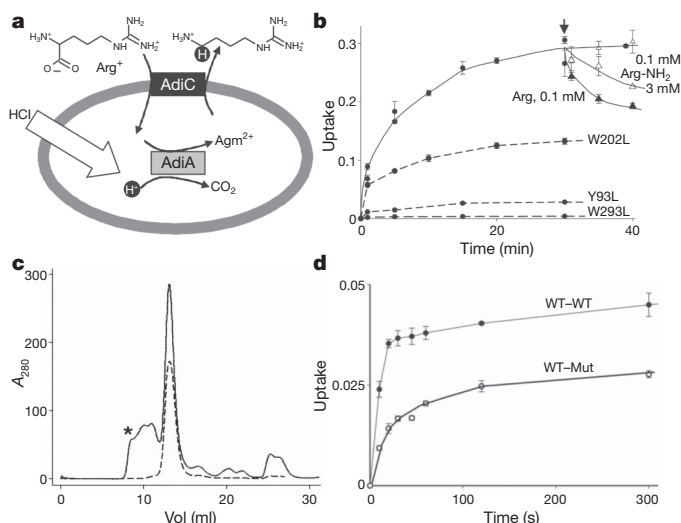
Figure 1 | AdiC physiology and function. a, Virtual proton pumping in extreme acid resistance. Schematic of Arg-dependent acid resistance in *E. coli*, with AdiC-mediated Arg/Agm exchange across the inner membrane coupled to acid-activated Arg-decarboxylase AdiA. Virtual proton is shown as a black circle. **b,** Selection of α-carboxylate for transport. Uptake (fraction of total counts added) of [^{14}C]Arg at 50 μM external concentration into AdiC-reconstituted liposomes (2 μg per mg lipid) was followed for 30 min, and then either Arg (filled triangles) or Arg-NH$_2$ (open triangles) was added (arrow) externally to the indicated concentration. Additional Arg uptake experiments used the indicated AdiC mutants (dashed curves). Error bars indicate s.e.m. of triplicate experiments. **c,** Proper assembly of tandem construct. Size-exclusion profiles of purified homodimeric AdiC (dashed trace) and WT–WT tandem (solid trace) in its final purification step immediately after elution from the co-affinity column. Material eluting between void volume of the Superdex 200 column (asterisk) and the main peak probably represents improperly assembled, oligomeric tandems. An identical profile is obtained with WT–Mut tandem in which the second subunit contains the W293L mutation (data not shown). **d,** A half-dead heterodimer is functionally active. Arg uptake time courses for WT–WT and WT–Mut tandems reconstituted at 0.2 μg per mg lipid, a low protein density where transporting liposomes carry only one copy of the reconstituted protein[6]. WT–WT transport rate is ~70% of normal homodimeric AdiC (data not shown). Error bars indicate s.e.m. of triplicate experiments.
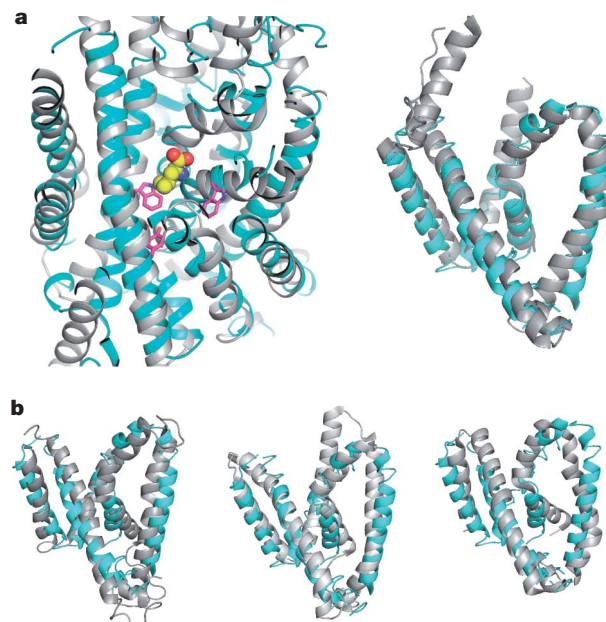
[1]Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA. [2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA.
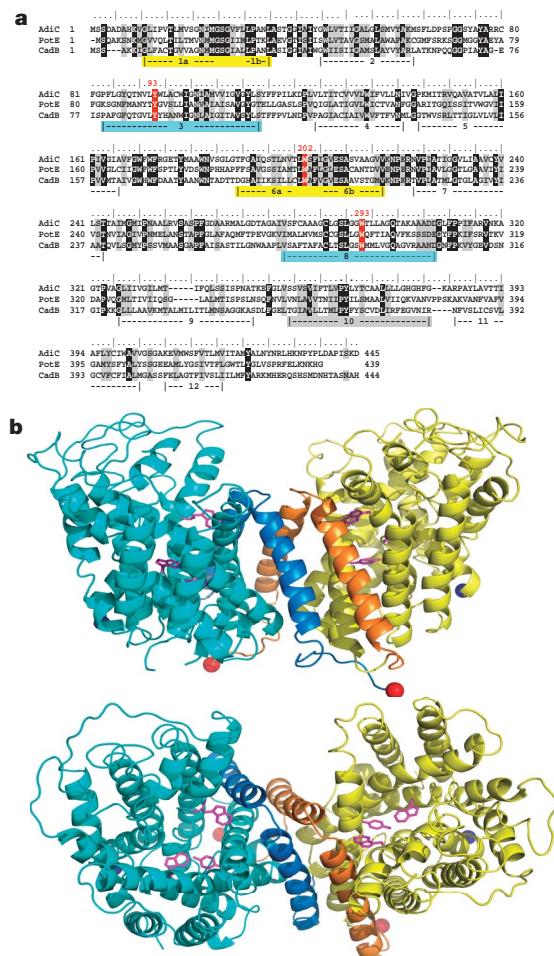
Figure 2 | Structure of AdiC. a, Sequence of AdiC aligned with two other *E. coli* virtual proton pump exchangers in the APC superfamily: PotE (ornithine/putrescine) and CadB (lysine/cadaverine). The five transmembrane helices lining the central cavity are highlighted, with matched colours representing helices paired by inverse repeat. Conserved aromatic residues discussed in the text are indicated in red. **b,** Ribbon diagrams of homodimer viewed from membrane, extracellular side up (top) and from extracellular solution (bottom), with Y93, W202 and W293 side chains indicated (pink sticks). N and C termini are indicated by blue and red spheres, respectively. TM11 and TM12 are indicated as a darker hue in each subunit.

Figure 3 | Structural alignment of AdiC with four transporter families. Alignments of AdiC (cyan) with four transporters (grey) are shown. **a,** LeuT, a wide-spectrum amino acid transporter. Left: TM1–TM10, showing bound substrate (space-filled) and conserved aromatics of AdiC (pink sticks). Right: TM1–TM5 core. **b,** Alignments of AdiC TM1–TM5 core with BetP (left), Mhp1 (centre) and SGLT1 (right). Cα r.m.s.d. with AdiC: LeuT (ref. 19), 3.1 Å; SGLT1 (ref. 20), 4.0 Å; Mhp1 (ref. 21), 3.4 Å; BetP (ref. 22), 3.6 Å.

substantial substrate specificity[6,7,14], the structural determinants of which are not known.

AdiC is a homodimer in detergent micelles and phospholipid membranes[6,7]. The fundamental question of whether subunit co-operation is required for transport or whether each subunit is a self-contained transporter must be settled before details of mechanism can be sensibly examined. We therefore designed tandem constructs containing two AdiC subunits joined together by a short linker. The 'WT–WT' tandem containing two wild-type subunits migrates identically to the wild-type homodimer on a size-exclusion column and shows similar transport activity (Fig. 1c), a result establishing that the tandem subunits fold and assemble normally. The transport-disruptive mutation W293L was then introduced into one of the subunits to form a 'WT–Mut' tandem for comparison of its transport activity to WT–WT. The result is clear (Fig. 1d): the WT–Mut tandem is functionally active, with an initial rate of Arg uptake roughly one-half of the WT–WT rate. Because the W293L substitution abolishes Arg binding[7], this experiment eliminates any transport model that requires substrate binding to both subunits during a single transport cycle, as in cooperative, half-of-sites mechanisms. The result implies that each subunit is itself a transporter, and that the

mechanistic underpinnings of substrate exchange are to be found within the subunit itself, as proposed previously on the basis of whole-cell studies for a homologous APC-superfamily protein[14].

A *Salmonella* AdiC homologue, 95% identical to the *E. coli* sequence, produced crystals diffracting anisotropically to 3.5 Å. Crystals of AdiC in a complex with a Fab fragment diffracted to 3.2 Å, and these, along with SeMet derivatives, were used for phasing (Supplementary Table 1). Serviceable crystals formed only in the absence of substrate. Experimental electron density maps of the Fab complex (Supplementary Fig. 1a) were sufficient for model building of the AdiC polypeptide chain, except for the first ten and last four residues, and for two disordered loops (residues 174–182, 316–321). Poly-Ala was built into a 22-residue extracellular loop linking transmembrane helices 5 and 6 and into several poorly ordered regions of the Fab fragments. The crystallographic arrangement of the complexes is unusual (Supplementary Fig. 1b), as the asymmetric unit contains two AdiC homodimers but only two Fab fragments, rather than four. This occurs because each Fab, by straddling the intracellular subunit interface near the two-fold axis of the homodimer axis, occludes the symmetry-related epitope. Most of the crystal contacts are mediated by Fab fragments, and only a few by feeble head-to-head encounters at neighbouring AdiC extracellular surfaces. A lower-resolution, molecular replacement structure without Fab closely recapitulates the transmembrane helices in the complex (Supplementary Fig. 1b), and thereby rules out problematic structural distortion by the Fab.

The roughly barrel-shaped AdiC subunit of ~45 Å diameter consists of 12 transmembrane helices (Fig. 2), transmembrane helix 1 (TM1) and TM6 being interrupted by short non-helical stretches in the middle of their transmembrane spans. Biochemical analysis of homologues places the amino and carboxy termini on the intracellular side of the membrane[15,16]. TM1–TM10 surround a large cavity exposed to the extracellular solution. These ten helices reprise in AdiC the remarkable inverted structural repeats in membrane protein families as functionally disparate as water/glycerol channels, H^+-coupled Cl^- antiporters, and Na^+-coupled symporters for a

variety of bio-organic compounds[17–22]. TM1–TM5 of AdiC align well with TM6–TM10 turned 'upside down' around a pseudo-two-fold axis nearly parallel to the membrane plane (α-carbon root mean squared deviation (r.m.s.d.) ~3.4 Å, Supplementary Fig. 2); thus, TM1 pairs with TM6, TM2 with TM7, and so on, but no hint of this alignment appears in the primary sequence. Helices TM11 and TM12, non-participants in this repeat, provide most of the 2,500 Å² homodimeric interface. Moreover, AdiC mirrors the common fold observed unexpectedly in four phylogenetically unrelated families of Na^+-coupled solute transporters[19–23]. This result, shown (Fig. 3a) by a structural alignment with the amino acid transporter LeuT confirms a recent prediction based on hydropathy analysis of APC-superfamily proteins[24]. Beyond the impressive match of the transmembrane helices, the alignment also shows substrate bound in LeuT coincident with the region of AdiC where the functionally critical residues Y93 and W293 reside. As a Na^+-independent antiporter joining the cluster of structurally similar families of Na^+-coupled symporters (Fig. 3b), AdiC further highlights emerging questions of convergent evolution versus deep links among ostensibly unrelated membrane proteins.

The central cavity stands out as a prominent feature of the structure (Fig. 4). This extracellular-facing aqueous cavern, ~25 Å wide at the rim, tapers to a floor situated in the centre of the protein about halfway through the membrane. The floor is formed by a pair of aromatic side chains, Y93 and W293, projecting inwards from TM3 and TM8, the long, tilted helices paired by the inverted repeat domains. An additional aromatic residue, W202, hangs 10–15 Å away on the cavity wall. These conserved residues have been variously proposed to contribute to substrate binding and transport in APC-type virtual proton pumps[7,14,25], as indicated above for AdiC (Fig. 1b). The wall, otherwise festooned with hydrophobic and polar moieties, is completely devoid of charged side chains. The cavity is unambiguously cut off from the cytoplasmic solution by tightly packed protein, ~15 Å thick.

The operation of any membrane transporter relies on a cycle of distinct protein conformations that expose substrate-binding sites alternately to the cytoplasmic and extracellular solutions and may additionally use intermediate 'occluded' states with substrates buried. Transport mechanisms are defined by rules linking substrate binding to transitions among these conformations. Because X-ray structures from the APC superfamily have not previously been described, the present view of AdiC in a single, substrate-free conformation is inadequate for laying out a mechanistic framework for this family of transporters. But the outward-facing structure here identifies a probable location for substrate binding at the cavity's floor. This suggestion is bolstered by three lines of argument. First, the narrowest part of a wide vestibule is a general expectation for the transport site in a coupled transporter[26]. Second, mutation of aromatic residues at this site inhibits transport activity in AdiC and homologues[7,14,25]. Finally, bound substrate in structurally aligned LeuT is found in this region of AdiC (Fig. 3a).

Viewed as a binding site for extracellular Arg, the cavity raises questions of molecular recognition underlying the protein's essential proton-extrusion function. Because Arg, Agm and analogues with similarly disposed charged groups are transported by AdiC (refs 6,7), the protein might be expected to offer oppositely charged residues for salt-bridge stabilization of substrates. But AdiC's physiological role in acid resistance quashes this expectation, as in a hydrated region exposed to pH 2, Glu/Asp side chains would be fully protonated and incapable of forming coulombic contacts. Indeed, charged residues are conspicuously absent from the cavity. The structure thus suggests, and we propose, that the outward-facing binding site neatly solves its electrostatic problem with aromatic side chains, which stabilize substrates with cation–π interactions[27,28]. A similar 'aromatic box' was recently observed at the binding site of a Na^+/betaine symporter[22]. However, it is clear that W202 is too far above the cavity's floor for Arg α-amino and γ-guanidino groups to contact all of these aromatic side chains simultaneously, and that the aromatics cannot by themselves mediate the nuanced substrate specificities of APC-type virtual proton pumps. We also note several backbone carbonyl oxygens on the repeat-related, non-helical stretches of TM1 and TM6, which might help stabilize substrates via any of the five H-bond donors of substrate guanidino groups. Resolution of the crucial biological issue—selectivity for the rare, negatively charged
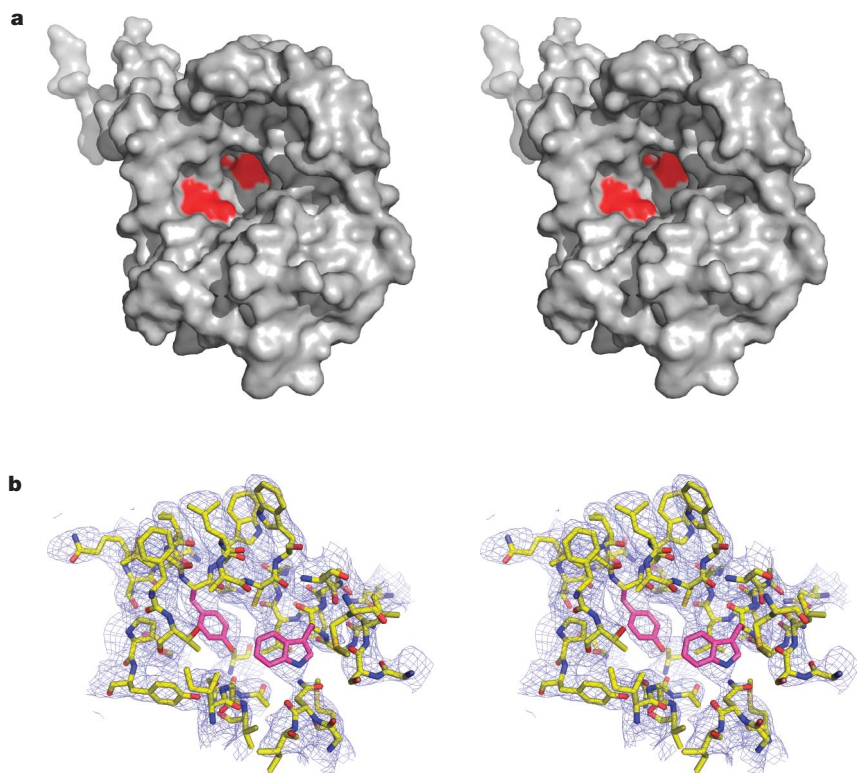
Figure 4 | AdiC cavity in outward-open conformation. a, Oblique stereo view from the extracellular side of a single AdiC subunit, showing locations of aromatic side chains (red). **b**, Close-up view of proposed substrate-binding region, with $2F_o - F_c$ map contoured at 1.5σ. Y93 and W293 are shown as pink sticks.

carboxylate of extracellular Arg—will require crystals with bound substrate, which we have so far failed to obtain by co-crystallization or soaking.

An alternative possibility exists, however, which if true would nullify all of the above ruminations on substrate selectivity. The structure here, while certainly outward-open, might not represent an Arg-binding form of the transport cycle at all, but rather an Agm-expelling form, the biological purpose of which is to efficiently rid the protein of substrate before beginning a new transport cycle. In other words, two different outward-open conformations, of high and low substrate affinities, might operate in transport; a similar situation might also apply to inward-facing forms. 'Dual-open' mechanisms like this have not been seriously considered for antiporters, but they are not *a priori* implausible. Indeed, such a mechanism would be well suited to the logic of virtual proton pumping, whereby Arg and Agm both move thermodynamically downhill, binding at high concentration and dissociating at low concentration, with gradients maintained by Arg decarboxylation. This structure provides no information about inward-open forms of AdiC, but a recent projection structure of AdiC in two-dimensional membrane crystals[7] could represent such a conformation, as it differs markedly from the same projection of our X-ray structure (Supplementary Fig. 3). Resolution of issues like this must await structures of AdiC under different conditions with substrates, and of other members of the APC superfamily.

Note added in proof: During the review of this paper, a 3.6-Å resolution structure of AdiC was reported[29]; the structure differs from the one described here in a register shift of 3–4 amino acid residues in transmembrane helices 6, 7 and 8.

METHODS SUMMARY

AdiC was expressed in *E. coli*, purified and reconstituted in liposomes[6]. AdiC function was assessed using the *E. coli* homologue by $[^{14}C]$Arg–Arg exchange (5 mM Arg inside, 50 μM $[^{14}C]$Arg outside), at protein density 0.2–2 μg AdiC per mg lipid[6]. Two tandem constructs used a 6-residue linker (GSAGGT) connecting the C terminus of the first AdiC subunit to the N terminus of the second. Monoclonal antibodies were raised by inoculating mice with *E. coli* AdiC and screening enzyme-linked immunosorbent assay (ELISA)-positive hybridomas for stable complexes by size-exclusion chromatography, crystallization and diffraction quality. Approximately 25 monoclonal antibodies were tested for crystallization to obtain Fab fragment 21 used here, which was derived from a type-2a IgG.

Salmonella enterica serovar Typhimurium AdiC in complex with Fab21 was purified on Superdex 200 in 100 mM NaCl, 5 mM decylmaltoside, 20 mM Tris-HCl pH 8, concentrated to 8 mg ml^{-1}, mixed with an equal volume of 30–35% (w/v) PEG 400, 100–200 mM $CaCl_2$, 100 mM glycine pH 9–9.5, and crystallized in hanging drops at 20 °C. Crystals were frozen after 2–4 weeks, and data sets were collected at NSLS, APS and ALS. SeMet derivatives were similarly treated. To increase redundancy of the anomalous signal in the *P*1 space group, data sets were collected with 4–6 360° passes. Experimental phases determined to 3.5 Å resolution by anomalous dispersion from two SeMet crystals—one at one wavelength and the other at two wavelengths—were combined using Sigmaa[30] and extended to 3.2 Å by noncrystallographic symmetry averaging (eight-fold for AdiC in the two crystal forms and two-fold for Fab21) and solvent flattening. Multi-domain, multi-crystal averaging using Dmmulti[31] greatly improved the phases. Sharpening of the data by a temperature factor of −80 Å^2 significantly enhanced the details of the electron density maps. Anomalous difference density maps identified 14 of the 15 Se atoms expected per subunit. Attempts to observe substrate density by soaking crystals in 5 mM substrates were unsuccessful.

1. Audia, J. P., Webb, C. C. & Foster, J. W. Breaking through the acid barrier: an orchestrated response to proton stress by enteric bacteria. *Int. J. Med. Microbiol.* **291**, 97–106 (2001).
2. Iyer, R. *et al.* A biological role for prokaryotic ClC chloride channels. *Nature* **419**, 715–718 (2002).
3. Iyer, R., Williams, C. & Miller, C. Arginine-agmatine antiporter in extreme acid resistance in *Escherichia coli. J. Bacteriol.* **185**, 6556–6561 (2003).
4. Gong, S., Richard, H. & Foster, J. W. YjdE (AdiC) is the arginine:agmatine antiporter essential for arginine-dependent acid resistance in *Escherichia coli. J. Bacteriol.* **185**, 4402–4409 (2003).
5. Foster, J. W. *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nature Rev. Microbiol.* **2**, 898–907 (2004).
6. Fang, Y., Kolmakova-Partensky, L. & Miller, C. A bacterial arginine-agmatine exchange transporter involved in extreme acid resistance. *J. Biol. Chem.* **282**, 176–182 (2007).
7. Casagrande, F. *et al.* Projection structure of a member of the amino acid/polyamine/organocation transporter superfamily. *J. Biol. Chem.* **283**, 33240–33248 (2008).
8. Maloney, P. C. Bacterial transporters. *Curr. Opin. Cell Biol.* **6**, 571–582 (1994).
9. Jack, D. L., Paulsen, I. T. & Saier, M. H. The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology* **146**, 1797–1814 (2000).
10. Nicholson, B. *et al.* Sustained nitric oxide production in macrophages requires the arginine transporter CAT2. *J. Biol. Chem.* **276**, 15881–15885 (2001).
11. Smith, P. A. *et al.* Electrogenic arginine transport mediates stimulus-secretion coupling in mouse pancreatic beta-cells. *J. Physiol. (Lond.)* **499**, 625–635 (1997).
12. Fuchs, B. C. & Bode, B. P. Amino acid transporters ASCT2 and LAT1 in cancer: partners in crime? *Semin. Cancer Biol.* **15**, 254–266 (2005).
13. Kaira, K. *et al.* L-type amino acid transporter 1 and CD98 expression in primary and metastatic sites of human neoplasms. *Cancer Sci.* **99**, 2380–2386 (2008).
14. Kashiwagi, K. *et al.* Identification of the putrescine recognition site on polyamine transport protein PotE. *J. Biol. Chem.* **275**, 36007–36012 (2000).
15. Kashiwagi, K. *et al.* Excretion and uptake of putrescine by the PotE protein in *Escherichia coli. J. Biol. Chem.* **272**, 6318–6323 (1997).
16. Hu, L. A. & King, S. C. Membrane topology of the *Escherichia coli* γ-aminobutyrate transporter: implications on the topography and mechanism of prokaryotic and eukaryotic transporters from the APC superfamily. *Biochem. J.* **336**, 69–76 (1998).
17. Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
18. Dutzler, R. *et al.* X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294 (2002).
19. Yamashita, A. *et al.* Crystal structure of a bacterial homologue of Na$^+$/Cl$^-$-dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
20. Faham, S. *et al.* The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na$^+$/sugar symport. *Science* **321**, 810–814 (2008).
21. Weyand, S. *et al.* Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322**, 709–713 (2008).
22. Ressl, S. *et al.* Molecular basis of transport and regulation in the Na$^+$/betaine symporter BetP. *Nature* **458**, 47–52 (2009).
23. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347–355 (2009).
24. Lolkema, J. & Slotboom, D.-J. The major amino acid transporter superfamily has a similar core structure as Na$^+$-galactose and Na$^+$-leucine transporters. *Mol. Membr. Biol.* **25**, 567–570 (2008).
25. Soksawatmaekhin, W. *et al.* Identification of the cadaverine recognition site on the cadaverine-lysine antiporter CadB. *J. Biol. Chem.* **281**, 29213–29220 (2006).
26. Guan, L. & Kaback, H. R. Lessons from lactose permease. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 67–91 (2006).
27. Burley, S. K. & Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **203**, 139–143 (1986).
28. Zacharias, N. & Dougherty, D. A. Cation-π interactions in ligand recognition and catalysis. *Trends Pharmacol. Sci.* **23**, 281–287 (2002).
29. Gao, X. *et al.* Structure and mechanism of an amino acid antiporter. *Science* **324**, 1565–1568 (2009).
30. Read, R. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
31. Cowtan, K. D. & Main, P. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr. D* **52**, 43–48 (1996).

Author Contributions Experiments were carried out and diffraction data collected by Y.F., H.J., T.S., L.K.-P., F.W., C.W. and C.M. Data were analysed by Y.F., H.J., Y.X. and C.M. The manuscript was prepared by Y.F., H.J., Y.X. and C.M.

Author Information Coordinates and structure factors of the AdiC–Fab complex have been deposited in the Protein Data Bank under accession number 3HQK. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.M. (cmiller@brandeis.edu).

ERRATUM

Kinematic variables and water transport control the formation and location of arc volcanoes

T. L. Grove, C. B. Till, E. Lev, N. Chatterjee & E. Médard

Nature **459**, 694–697 (2009)

In this Letter, two of the Table 1 headings were incorrectly reversed. The Central Aleutians in ref. 13 and Western Aleutian in ref. 14 are comparable geographic regions. This discrepancy was not noted in the original Table 1. The corrected Table is printed below.

Table 1 | Outputs from our geodynamic models with measured arc volcano locations

Arc name	Convergence rate (km per million years)	Slab dip (degrees)	Arc–trench distance (km)	Depth to slab (km)
Central Aleutians	60	58	176 (ref. 14); 44 (ref. 13)	99 (ref. 14); 90 (ref. 13)
Model similar to the central Aleutians	60	64	40	80
New Zealand	36	50	96 (ref. 14); 92 (ref. 13)	123 (ref. 14); 110 (ref. 13)
Model similar to New Zealand	40	40	104	70

Data for modern subduction zones from ref. 13 did not contain direct measurements of distance from trench to arc. We calculate this value from their measurements of depth to slab and the slab dip by a simple geometric approximation: (depth to slab)/tan (slab dip).

CORRIGENDUM

Generation of pluripotent stem cells from adult human testis

Sabine Conrad, Markus Renninger, Jörg Hennenlotter, Tina Wiesner, Lothar Just, Michael Bonin, Wilhelm Aicher, Hans-Jörg Bühring, Ulrich Mattheus, Andreas Mack, Hans-Joachim Wagner, Stephen Minger, Matthias Matzkies, Michael Reppel, Jürgen Hescheler, Karl-Dietrich Sievert, Arnulf Stenzl & Thomas Skutella

In this Letter, we omitted to disclose that the original patient consent forms to collect the material used to derive the pluripotent stem cells precluded distribution to third parties, as ruled on 8 December 2008, by the Ethics Commission of the School of Medicine and University Hospital Tübingen. Moreover, the Commission also ruled that the materials used to generate the lines were obtained from individuals who had signed consent forms that did not allow retention of the cells in culture for more than 3 years. We have now received broader consent of a few individuals permitting distribution of cells. These cells will be cultivated, and after quality tests they will be distributed to other scientists. The other individuals whose tissues were used in the study have opted for the terms of the initial consent, and thus their cell lines will be destroyed when the 3-year period runs out. We are now using the described protocols to generate new cell lines from individuals who have given explicit consent for distribution of lines to scientists.

CORRIGENDUM

Dual nature of the adaptive immune system in lampreys

Peng Guo, Masayuki Hirano, Brantley R. Herrin, Jianxu Li, Cuiling Yu, Andrea Sadlonova & Max D. Cooper

In this Letter, the scale bars in Fig. 5c are incorrectly listed as 1 mm, the correct size of the scale bars is 1 μm.

Q&A

Lucia Votano is the incoming director of the National Laboratory in Gran Sasso, Italy, one of four national labs in the country's National Institute of Nuclear Physics.

What does your new post mean to you?

My appointment, which begins in September, represents the first time a woman has been named as director of any of the National Institute's four labs. The National Laboratory is the world's largest underground lab for astroparticle physics, and my election is a sign that the institute is taking a big step forward and that women in science are now considered as qualified as men for such a post.

On a personal level, the position is a natural progression from the beginning of my career as a physicist. It will be my last role — a fitting completion. After this I will retire.

What do you consider your biggest challenge?

In April, a devastating earthquake in our region killed more than 200 people, injured 1,500 and left 55,000 homeless. Our lab activity never stopped completely because our building was built to withstand seismic activity. But people who work and live here are homeless. For me, the important challenge is helping the people. We have donated space for the region to set up a school, we host municipal and other meetings and we are involved in setting up a new interactive science museum in the ruined centre of L'Aquila.

What's your greatest scientific achievement?

I'm quite proud of my present experiment, the Oscillation Project with Emulsion Tracking Apparatus (OPERA), in which a muon-neutrino beam generated at the CERN particle-physics lab near Geneva, Switzerland, is directed towards our detectors. I chaired the OPERA policy- and decision-making board and helped to determine the design, construction and definition of the beam apparatus. We are trying to prove neutrino oscillation, in which a muon, tau or electron neutrino changes from one type to another.

Do bias and bureaucracy plague physics in Italy?

In particle physics, we may be in a better position than other fields. Politics has always been far, far away from our rules and internal government. Perhaps it's because we are always involved in international collaborations, so we must be able to work well with other scientists and other institutes. Otherwise we would disappear from the experiment.

What is the secret of scientific success?

You must do your job with great devotion, dedication, diligence and care. You must continually compare your ideas with those of your colleagues and share your objectives with colleagues. You do not impose your ideas on anyone.

What do you value most about the scientific process?

Intellectual honesty. It's a value that we should try to export to other fields, such as politics. ∎

Interview by Karen Kaplan

POSTDOC JOURNAL
Failure to communicate

Sometimes in science we fail to communicate effectively. As I try to tell colleagues and friends what I do in the lab, I'm reminded of a germane quote. In the film *Cool Hand Luke*, Luke refuses to submit to the system in a Florida prison camp. After many raucous encounters with Luke, the outraged prison warden concludes: "What we've got here is failure to communicate."

I talk about science with colleagues, friends and family every day. Science can be regarded as a separate language, replete with its own vocabulary, syntax and usage. When I share this language with colleagues, no translation is needed. But with friends and family, I have to translate to be fully understood.

Take my PhD thesis. When co-workers asked what the title would be, I could say: "High-resolution genome-wide mapping of the yeast transcription machinery".
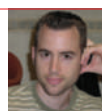
Implicit in this is my use of a method of chromatin immunoprecipitation followed by hybridization to microarrays or high-throughput sequencing for "high-resolution genome-wide mapping".

But when my family asked the same question, I had to say, "I study how your cells control what genes are turned on and off."

Having found it's not easy to explain my research in either the untranslated or translated version, I've learned a valuable lesson. Regardless of the intended audience, being clear and concise pays off if you want to be understood. ∎

Bryan Venters is a postdoctoral fellow at the Center for Eukaryotic Gene Regulation at Pennsylvania State University, University Park.

IN BRIEF

UK lab to get new digs

The laboratory at the centre of the UK foot-and-mouth disease outbreak in 2007 is getting a £100-million (US$165-million) renovation aimed at attracting scientists back to the Institute for Animal Health (IAH).

The new facilities at the Pirbright Laboratory, where the IAH studies how to control and treat infectious diseases in animals, are due for completion in 2013.
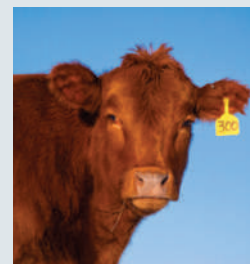
The lab, shared by the IAH and vaccine manufacturer Merial, was linked to two foot-and-mouth outbreaks (see *Nature* 449, 259–260; 2007). Reports both before and afterwards called the lab "shabby", "substandard" and "dilapidated".

Revamping it and improving biosecurity will make it easier to recruit and retain quality scientists, says science minister Paul Drayson. The renovation is being funded by the new Department for Business, Innovation and Skills.

Banking on energy

Research scientists are set to benefit from a $3-billion pledge by the US Department of Energy for alternative-energy research and manufacturing. It will help fund such areas as battery manufacturing for hybrid and electric vehicles; new energy research centres; and upgrades of energy-research infrastructure at national labs.

The largest sum, $2.4 billion, is expected to support research at vehicle and battery manufacturers (see *Nature* 460, 791; 2009). Another $377 million will support 46 'energy frontier' centres at universities, national labs and research institutions. These will focus on basic and applied science in various fields including biofuels, solar cells and carbon sequestration (see *Nature* 459, 285; 2009).

Also, $327 million will pay for research, instrumentation and lab infrastructure at 10 national labs. Of that, $60 million, the largest allocation, goes to Fermilab in Batavia, Illinois, for a particle accelerator.

The funding, announced earlier this month, is part of President Barack Obama's economic recovery plan.

S. WIRZBA/CORBIS

The Piltdown angel

Dig for victory.

John P. Boyd

Our archaeological team had found the skeleton of a small, six-winged hominid in an aluminium casket buried under an Irish peat bog. But I didn't suspect Andrei until one night when he pointed at Mars.

"The colonists will envy our Angel." He laughed. "If they can't dig up his brother, they'll make one!" He sighed. "Will the World Space Agency really abandon Mars, Adelard?"

I shrugged. The Climate War is turning us inward. Bangladesh is underwater. Florida's an archipelago. Waves of refugees are moving, all over the globe. Every country feels like Hitler in the Bunker. Lost souls, just trying to survive. "Maybe in 50 years there'll be time for dreams, Andrei — and Mars."

I couldn't sleep. Andrei's smirk, so like Dad's dogsbody …

My father had once been a Harvard dean. He named me after the scholar Adelard of Bath, who had travelled to Antioch to consult original sources around 1115. But Dad was too trusting. He kept churning out books as dean, a mere proofreader of tomes ghost-written by his assistants. One lowly postdoc, suffering from both depression and anxiety disorder, scribbled stories half-remembered from undergraduate classes instead of consulting the archives. My father was only suspended without pay for a year, losing his deanship and chair.
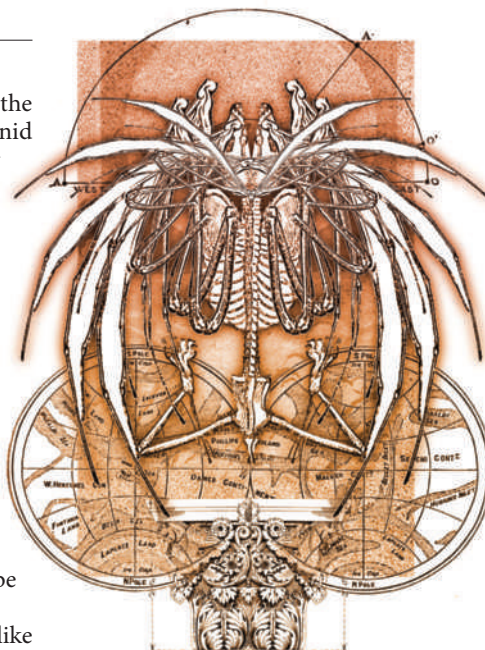
I would not be destroyed like Dad! Sir Keith gave me a way to fight back a few hours after we pried open the casket. He said: "This is now a crime scene. We've found something big. But radical discoveries are accepted only with the strongest, most indisputable evidence. The dig protocols must be followed to the letter as if each was a test by a malevolent deity on which your life depends. Because this *is* a test. We'll be strongly criticized, examined and probed, looking for the slightest weakness. Jealousy is a chronic scholarly virus!"

Andrei hated the fussiness. "Like a Soviet submarine!" he grumbled. "No man can move a pebble without another beside him, like his political commissar."

Sarah Frane was a very good commissar. Unlike Andrei, a soil geologist, she was soft science, and believed in history the way a medieval nun believed in Mary.

When the heavy rains came, her allergies worsened. I teased her.

"Did you ever think, Sarah, that our Angel might be another Piltdown Man?"

She scowled.

"An early hominid excavated in East Sussex in 1912. Orang-utan jaw mated with a modern skull, age-stained with iron and chromic acid."

She shook her head. "I'm busy!"

An hour later, I added: "The teeth were filed for consistency with the prejudice that big brains evolved before an omnivorous diet." I blocked her view of the soil-sifter. "By 1915, Boule and Miller independently wrote that the jaw was an ape's."

She counterattacked. "You're our palaeontologist. You've touched the bones. X-ray, CT scan, chemical analysis."

A riposte! "They're not bird or pterodactyl wings," I admitted. "They're real bones, like nothing else on Earth."

Still, she was worried, and I attacked after lunch. "Why would aliens hide from us, and then foolishly bury one of their own where it might be dug up?"

She looked like she was going to cry. A few minutes later, I said: "In 1923, Weidenreich was spot on. It still took 30 years for the scientific community to accept he was right."

She ran off just to get away from me, then scuttled back under the tent with fogged glasses and face streaked with rain.

"Let me towel —"

She fled to the main tent to clean up.

I was alone only for minutes, but that was plenty of time to contaminate the Angel's strata with a 40-year-old American quarter, a gum wrapper and a paper clip.

Later, while an undergraduate was Sarah's 'political commissar', I heard a little shriek and knew my anachronisms had been found.

Every day, I waited for the chaos and recriminations of my senior year in high school when I suddenly became the son of a criminal. Nothing happened. Then Sir Keith and Mary Manon announced our 'Angel', and I pretended to smile for the scanners.

When we walked home under the stars, Andrei said: "You are very quiet, amigo."

"Sir Keith and M-Squared will get the glory. Book tours, honorary degrees. A life peerage. We're just shovel-guys."

"Eh. Not shovel-guy. You're one of the world's best bone-men." He smiled. "But you *will* get tenure. And I *will* do stratigraphy on Mars!"

I said angrily: "That's what this has all been about. Saving Mars! Diverting funds from history and archaeology and understanding ourselves to chasing phantasms on Pluto. And it's all a *hoax*."

Andrei frowned.

I shouted: "Sarah found contamination! Modern artefacts in the same layer as the Angel."

He waved his hand dismissively. "*Every* site is contaminated. Every dig team drops a penny in the dig, or a food wrapper. Every historical event has anomalies, spawning dozens of conspiracy theories. Mutually exclusive theories. "We talked to Sarah. Our mousy little friend was upset, but she wanted to believe. She did believe."

That night, I lay on the straw for hours. Andrei could jigger the soil analysis, but who had forged the six-winged skeleton? I had touched the curving wings; I knew my bones. Impossible to fake.

There was compelling evidence that the angel was a hoax. There was compelling evidence it was not. It was thus in all scientific revolutions, both true and failed. There was in each a time of wonder and confusion when logic was inadequate and only faith remained, dignified as intuition.

I walked restlessly until dawn and shouted at the stars: "Little angel, where are you *from*?"

What did I believe? And would I ever *know*? ∎

John P. Boyd has published 16 previous stories; this is his second for *Nature*. He is professor of atmospheric and oceanic science at the University of Michigan. Join the discussion of Futures in *Nature* at http://tinyurl.com/kkh3kt

JACEY